

DOCUMENT RESUME

ED 243 892

TM 830 020

AUTHOR Wainer, Howard
 TITLE How to Display Data Badly. Technical Report No. 82-33.
 INSTITUTION Educational Testing Service, Princeton, NJ. Program Statistics Research Project.
 REPORT NO ETS-RR-82-35
 PUB DATE Aug 82
 NOTE 68p.; Paper presented at the Annual Meeting of the American Statistical Association (Cincinnati, OH, August 16-19, 1982).
 AVAILABLE FROM Educational Testing Service, Research Publications R-116, Princeton, NJ 08541.
 PUB TYPE Speeches/Conference Papers (150) -- Guides - Non-Classroom Use (055) -- Reports - Research/Technical (143)
 EDRS PRICE MF01/PC03 Plus Postage.
 DESCRIPTORS Charts; Communication (Thought Transfer); *Graphs; *Statistical Data; Tables (Data)
 IDENTIFIERS *Data Display

ABSTRACT

The aim of good data graphics is to display data accurately and clearly. This definition is used as a point of departure for developing 12 rules of bad data display: (1) show as little data as possible (minimize the data density); (2) hide what data you do show (minimize the data-ink ratio); (3) ignore the visual metaphor altogether; (4) only order matters (The Pravda School of Ordinal Graphics); (5) graph data out of context; (6) change scales in mid-axis; (7) emphasize the trivial (ignore the important); (8) jiggle the baseline; (9) order graphs and tables alphabetically; (10) label illegibly, incompletely, incorrectly, and ambiguously; (11) more is murkier (more decimal places and more dimensions); and (12) if it has been done well in the past, think of a new way to do it. Although the tone of this presentation is light, and points in the wrong direction, the aim is serious. The 12 "rules" point clearly toward an outlook that provides many hints for good display. (BW)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

How To Display Data Badly

Howard Wainer

Program Statistics Research
Technical Report No. 82-33

Research Report RR-82-35

August 1982

This is the text of an invited address to the American Statistical Association. It was supported in part by the Program Statistics Research Project of the Educational Testing Service.

Abstract

Methods for displaying data badly have been under development for many years, and a wide variety of interesting and inventive schemes have emerged. Presented here is a synthesis yielding the twelve most powerful techniques that seem to underlie many of the realizations found in the literature. These twelve (the dirty dozen) are identified and illustrated.

Table of Contents

	<u>Page</u>
Introduction	1
Showing Data	1
Showing Data Accurately	5
Showing Data Clearly	8
Two Other Techniques That Need Not Lead Us Astray, But Usually Do	14
Summing Up	15
Figure Captions	20
References	23

Introduction

The display of data is a topic near and dear to my heart, and one which has occupied the thoughts of many scholars for almost two hundred years. During this time there have been a number of attempts to codify standards of good practice (e.g., ASME Standards, 1915; Cox, 1976) as well as a number of books that have illustrated them (i.e., Bertin, 1973, 1977, 1981; Schmid, 1954; Schmid & Schmid, 1979; Tufte, 1982). The last decade or so has seen a tremendous increase in the development of new display techniques and tools, which have been reviewed recently (Wainer & Thissen, 1981). We wish to concentrate on methods of data display that leave the viewers as uninformed as they were prior to seeing the display or, worse, those which induce confusion. Although such techniques are broadly practiced, to my knowledge, they have not as yet been gathered into a single source, nor carefully categorized. This paper is meant as the beginning of such a synthesis.

The aim of good data graphics is to display data accurately and clearly. Let us use this definition as a point of departure for categorizing methods of bad data display. The definition has three parts. These are:

- 1) showing data
- 2) showing data accurately
- 3) showing data clearly

Thus if we wish to display data badly, we have three avenues to follow. Let us examine them in sequence, parse them into some of their component parts, and see if we can identify means for measuring the success of each strategy.

Showing Data

Obviously, if the aim of a good display includes conveying information, the less information carried in the display the worse it is. Tufte (1982)

has operationalized this into the data density index (ddi), which is 'the number of numbers plotted per square inch.' This easily calculated index is often surprisingly informative. In popular and technical media we have found a range from .1 to 362, with the data graphics in Pravda occupying the bottom of the list--a model for empty display. This provides us with the first rule of bad data display:

*Rule 1--Show as little data as possible (minimize the data density).

What does a data graphic with a ddi of .3 look like? Shown in Figure 1 is a graphic from the book Social Indicators III, done in four colors (original size 7" by 9") that contains 18 numbers ($18/63 = .3$). The median data graph in Social Indicators III has a data density of .6 numbers/in²;

Insert Figure 1 about here

this one is not an unusual choice. Shown in Figure 2 is a plot from the Journal of the American Statistical Association (JASA) with a ddi of .5 (it shows 4 numbers in 8 in²). This is unusual for JASA, where the median data graph has a ddi of 27. In defense of the producers of this plot, the point of the graph is to show that a method of analysis suggested by a critic of their paper was not fruitful. I suspect that prose would have worked pretty well also.

Insert Figure 2 about here

While arguments can be made that high data density does not imply that a graphic will be good, nor low density bad, it does reflect on the efficiency of the transmission of information. Obviously, if we hold clarity and accuracy constant, more information is better than less. One of the great assets of

graphical techniques is that they can convey large amounts of information in a small space. Of perhaps some interest, the only publication that we have found that has lower information density graphics than Social Indicators III is Pravda. Shown in Table 1 (adapted from Tufte, 1982) are some summary values for the data density of graphics in a variety of publications.

Insert Table 1 about here

Higher density graphs are easily found. Note in Figure 3 a plot of weather from The New York Times (January 11, 1981) that contains 2,222 numbers in 34 in² for a ddi of 65.

Insert Figure 3 about here

We note that when a graph contains little or no information the plot can look quite empty (Figure 2) and thus raise suspicions in the viewer that nothing is to be communicated. A way to avoid these suspicions is to fill up the plot with nondata figurations: what Tufte has termed "Chartjunk." In Figure 4 is a plot of the labor productivity of Japan relative to that of the United States. It contains one number for each of three years. Obviously, a graph of such sparse information would have a lot of blank space, so filling the space hides the paucity of information from the reader.

Insert Figure 4 about here

A convenient measure of the extent to which this practice is in use is Tufte's 'data-ink ratio.' This measure is the ratio of the amount of ink used to graph the data divided by the total amount of ink in the graph. The closer to zero this gets, the worse the graph.

The notion of the data-ink ratio brings us to the second principle of bad data display.

*Rule 2--Hide what data you do show (minimize the data-ink ratio).

One can hide data in a variety of ways. One method that occurs with some regularity is to hide the data in the grid. The grid is useful for plotting the points, but only rarely afterwards. Thus to display data badly, use a fine grid and plot the points dimly. This is exemplified by a plot in Kelly, Ayres and Bowen (1967), reproduced in Figure 5. The data-ink ratio for

Insert Figure 5 about here

this plot is close to zero. Is it possible to worsen it? The paper from which this plot was abstracted has been subsequently anthologized, and the editor, William J. Crotty (1970), hit upon a plan to decrease the data-ink ratio still further. He omitted the data entirely, emphasized the grid and still maintained the extraneous 105% border. He utilized a full page of a book to reproduce a piece of graph paper with a diagonal line. He also realized the ultimate--an absolute zero data-ink ratio!

Insert Figure 6 about here

Tufte (1970), plotting almost the residual of Crotty's plot from Kelly, Ayres and Bowen's, produced a graph with a data-ink ratio of about .7 (see Figure 7).

Insert Figure 7 about here

A second way to hide the data is in the scale. This corresponds to blowing up the scale (i.e., looking at the data from far away) so that any variation in the data is obscured by the magnitude of the scale. One can justify this practice by appealing to 'honesty requires that we start the scale at zero,' or other sorts of sophistry.

In Figure 8 is a plot that (from SI3) effectively hides the growth of private schools in the scale.

Insert Figures 8, 9 and 10 about here

A redrawing of the number of private schools on a different scale conveys the growth that took place during the mid-1950's (Figure 9). A more famous example is shown in Figure 10.

To conclude this section, we have seen that we can display data badly by either not including them (Rule 1) or by hiding them (Rule 2). We can measure the extent to which we are successful in excluding the data through the data density; we can sometimes convince viewers that we have included the data through the incorporation of chartjunk. Hiding the data can be done using an overabundance of chartjunk, or by cleverly choosing the scale so that the data disappear. A measure of the success we have achieved in hiding the data is through the data-ink ratio.

Showing Data Accurately

The essence of a graphic display is that a set of numbers having both magnitudes and an order are then represented by an appropriate visual metaphor-- the magnitude and order of the metaphorical representation matches the numbers. We can display data badly by ignoring or distorting this concept.

*Rule 3--Ignore the visual metaphor altogether.

If the data are ordered and if the visual metaphor has a natural order, a bad display will surely emerge if you shuffle the relationship. In Figure 11 note the bar labeled 14.1 is longer than the bar labeled 18. Another method is

Insert Figures 11, 12, 13, and 14 about here

to change the meaning of the metaphor in the middle of the plot. In Figure 12 the dark shading represents imports on one side and exports on the other. This is but one of the problems of this graph; more serious still is the change of scale. There is also a difference in the time scale, but that is minor. A redone version is shown in Figure 13. A common theme in Playfair's (1786) work was the difference in imports and exports. In Figure 14 is a 200-year-old graph that tells the story clearly. Two such plots would have illustrated the story surrounding this graph quite clearly.

*Rule 4--Only order matters (The Pravda School of Ordinal Graphics).

One trick often utilized is to use length as the visual metaphor when area is what is perceived. This was used quite effectively by the Washington Post in Figure 15. Note that this graph also has a low data density (.1), and its data-ink ratio is close to zero. We can also calculate Tufte's (1982) lie

Insert Figures 15 and 16 about here

factor for this graph. The lie factor in this instance is the perceived change in the value of the dollar from Eisenhower to Carter divided by the actual change. I read and measure thus:

<u>Actual</u>	<u>Measured</u>
$\frac{1.00 - .44}{.44} = 1.27$	$\frac{22.00 - 2.06}{2.06} = 9.68$

$$\text{lie factor} = 9.68/1.27 = 7.62$$

This distortion of over 700% is substantial, but by no means a record.

Much greater distortion can be accomplished by using volumes rather than areas, and by playing around with the scale and origin. In Figure 17 the volumetric increase in beer sales is a whopping 38,200%, whereas the actual increase is a respectable 29%, for a lie factor of 1317 (131,724%). This is a mark that even Pravda's artists have to respect. A more accurate depiction is shown in Figure 18. Figures 15 and 17 are examples of Tufte's (1976) "Goosing up the effect by squaring (cubing) the eyeball trick."

Insert Figures 17 and 18 about here

*Rule 5--Graph data out of context.

Often we can modify the perception of the graph (particularly for time series data) by choosing the interval displayed carefully. A precipitous drop can disappear by choosing a starting date just after the drop. (See Figures 19 and 20.) Similarly, we can turn slight meanders into sharp changes by focussing in on a single meander and expanding the scale. Often the choice of

Insert Figures 19 and 20 about here

scale is arbitrary but can have profound effects upon the perception of the display. Automatic rules do not always work, and wisdom is always required.

In this section we have discussed three rules for the accurate display of data. One can compromise accuracy by ignoring visual metaphors (Rule 3), by

only paying attention to the order of the numbers and not their magnitude (Rule 4), or by showing data out of context (Rule 5). We have advocated the use of Tufte's lie factor as a way of measuring the extent to which the accuracy of the data has been compromised by the display. One can think of modifications that would allow it to be applied in other situations, but we leave such expansion to other accounts.

Showing Data Clearly

In this section we discuss methods for displaying data badly that do not seem as serious as those described previously, i.e., the data are displayed, and they might even be accurate in their portrayal. Yet subtle (and not so subtle) techniques can be used to obscure effectively the most meaningful or interesting aspects of the data. It is more difficult to provide objective measures of presentational clarity, but we rely on the reader to judge from the examples presented.

*Rule 6--Change scales in mid-axis.

This is a powerful technique that can make large differences look small and make exponential changes look linear.

In Figure 21 is a graph that supports the associated story about the skyrocketing circulation of the New York Post relative to the plummeting Daily News circulation. The reason given is because New Yorkers 'trust' the Post.

Insert Figure 21 about here

It takes a careful look to note the 700,000 jump that the scale makes between the two lines.

In Figure 22 is a plot of physicians' incomes by time. It appears linear with a slight tapering off in recent years. A careful look at the scale shows

that it starts out plotting every eight years and ends up yearly. A more regular scale (in Figure 23) tells quite a different story.

Insert Figures 22 and 23 about here

*Rule 7--Emphasize the trivial (ignore the important).

Sometimes the data which are to be displayed have one aspect that is important and others that are trivial. The graph can be made worse by emphasizing the trivial part. In Figure 24 we have a page from SI3 that compares the income levels of men and women by educational levels. This reveals the not surprising result that better educated individuals are paid better than more poorly educated ones and that changes across time are modest. The comparison of greatest interest and current concern, comparing salaries between sexes within education level, must be made clumsily by vertically transposing from one graph to another. It seems clear that Rule 7 must have

Insert Figure 24 about here

been operating here, for it would have been easy to place the graphs side-by-side and allow the comparison of interest to be made more directly (Figure 25). Viewed strictly from a data analytic context, we note that there are two large main effects (education and sex) and a small time effect. This would have implied a plot that showed the large effects clearly and placed the smallish time trend into the background (Figure 26).

Insert Figures 25 and 26 about here

*Rule 8--Jiggle the baseline.

Making comparisons is always aided when the quantities being compared start from a common base. Thus we can always worsen the graph by starting from different bases. Such schemes as the hanging or suspended rootogram, and residual plot are meant to facilitate comparisons. In Figure 27 is a plot of U.S. imports of red meat taken from the Handbook of Agricultural Charts published by the U.S. Department of Agriculture. Shading beneath each line is a convention that indicates summation, telling us that the amount of each kind of meat is added to the amounts below it. Because of the dominance and the fluctuations of beef and veal, it is hard to see what the changes are in the other kinds of meat--Is the importation of pork increasing? Decreasing? Staying constant? The only purpose for stacking is to graphically indicate the total summation. This is easily done through the addition of another line for TOTAL. Note that this will always be clear and never intersect with the other lines on the plot. A version of these data with the separate amounts of each meat shown clearly as well as a summation line is shown in Figure 28. Note how easily one can see the structure of import of each kind of meat now that the standard of comparison is a straight line (the time axis) and no longer the import amount of those meats with greater volume.

Insert Figures 27 and 28 About Here

*Rule 9--Alabama First!

Ordering graphs and tables alphabetically can effectively obscure structure in the data that would have been obvious had the display been ordered by some aspect of the data. One can defend oneself against criticisms by pointing out that alphabetizing 'aids in finding entries of interest.' Of

course with lists of modest length, such aids are unnecessary; with longer lists the indexing schemes common in 19th Century statistical atlases, provide easy lookup capability.

Figure 29 is another graph from SI3 showing life expectancies for 10 industrialized nations divided by sex. The order of presentation is alphabetical (with the USSR positioned as Russia). The message we get is that there is little variation, and women live longer than men. Redone as a stem-and-leaf diagram (Figure 30 is simply a reordering of the data with spacing proportional to the numerical differences), the magnitude of the sex difference leaps out at us. We also note the USSR as an outlier for men.

Insert Figures 29 and 30 about here

Note in Figure 31 another plot from SI3 that has the same formal structure as Figure 29 except that the countries are ordered by their 1976 motor vehicle fatality rates. We easily see the profound differences (a factor of four between the U.S. and Belgium!) as well as the large advances made in auto safety over the seven years depicted. Last, the unusual gain made by Japan in auto safety is visible. We do not contend that a better display still could not be made, but the simple ordering has helped. This could have been obscured by alphabetizing. In tabular presentations the effects of alphabetizing (or presenting the data ordered by some other aspect unrelated to the data) can be particularly profound. This has been discussed eloquently by Ehrenberg (1977).

Insert Figure 31 about here

*Rule 10--Label: (a) illegibly, (b) incompletely, (c) incorrectly,
and (d) ambiguously.

There are many instances of labels that either do not tell the whole story, tell the wrong story, tell two or more stories, or are so small that one cannot figure out what story they are telling. A favorite example of 'small labels' is from The New York Times (August 1978), in which the article complains that fare cuts lower commission payments to travel agents. The graph (Figure 32) supports this view until one notices the tiny label indicating that the small bar showing the decline is for just the first half of 1978. This omits such heavy travel periods as Labor Day, Thanksgiving, Christmas, etc., so that merely doubling this is probably not enough. Nevertheless, when this bar is doubled (Figure 33), we see that the agents are doing very well indeed, relative to earlier years.

Insert Figures 32 and 33 about here

*Rule 11--More is murkier: (a) more decimal places and (b) more dimensions.

We often see tables in which the number of decimal places presented are far beyond the number that can be perceived by a reader. They are also commonly presented to more accuracy than is justified. A display can be made clearer by presenting less. In Table 2 is a section of a table from Dhariyal and Dudewicz's 1981 JASA paper. The table entries are presented to 5 decimal places! In Table 3 is a heavily rounded version that shows what the authors intended clearly. It also shows that the various columns might have a substantial redundancy in them (the maximum expected gain with $b/c = 10$ is about 1/10th that with $b/c = 100$ and 1/100th that of $b/c = 1000$). If this is true, the entire table could have been reduced substantially.

Insert Tables 2 and 3 about here

Just as increasing the number of decimal places can make a table harder to understand, so can increasing the number of dimensions make a graph more confusing. We have already seen how extra dimensions can cause ambiguity (is it length or area or volume?). In addition, human perception of areas is inconsistent. Just what is confusing and what is not is sometimes only a conjecture, yet a hint that a particular configuration will be confusing is obtained if that display confused the grapher. Shown in Figure 34 is a plot of per share earnings and dividends over a six-year period. We note (with some amusement) that 1975 is the side of a bar--the third dimension of this bar (rectangular parallelepiped?) chart has confused the artist!! I suspect that 1975 is really what is labeled 1976, and the unlabeled bar at the end is probably 1977. A simple line chart with this interpretation is shown in Figure 35.

Insert Figures 34 and 35 about here

In this section we have illustrated six more rules for displaying data badly. These rules fall broadly under the heading of how to obscure the data. The techniques mentioned were: change the scale in mid-axis, emphasize the trivial, jiggle the baseline, order the chart by a characteristic unrelated to the data, label poorly, and include more dimensions and/or decimal places than are justified or needed. Each of these methods will work in combination with the others to produce graphs and tables of little use. Their common effect will usually be to leave us uninformed as to the points of interest in the data, although sometimes they will misinform us; the physicians' income plot in Figure 22 is a prime example.

Two Other Techniques That Need Not Lead Us Astray, But Usually Do

Use color. The availability of color usually means that there are additional parameters that can be misused. I refrain from presenting examples, for that will effectively prevent the subsequent publication of this paper, although Census's Two-Variable Color Map is a wonderful example of how color in a graph can seduce us into thinking that we are communicating more than we are. This leads us to the last rule.

*Rule 12--If it has been done well in the past, think of a new way to do it.

The two-variable color map was done rather well by Mayr 100 years before the Census version. He used bars of varying width and frequency to accomplish gracefully what Census used varying saturations to do clumsily.

A particularly enlightening experience is to look carefully through the six books of graphs that William Playfair published during the period 1786-1822. One discovers clear, accurate and data-laden graphs utilizing many ideas that are useful and too rarely applied today. In the course of preparing this paper I spent many hours looking at a variety of attempts to display data. Some of the horrors that have been presented were the fruits of that search. In addition, jewels sometimes emerged. I have saved the best for last, and would like to conclude with one of those jewels--my nominee for the title of "World's Champion Graph." It was produced by Minard in 1824 and shows six variables important to the understanding of the events involved in the French Army's march into Russia in 1812-1813. This graph (originally in color) is reproduced in Figure 36. The width of the river is proportional to the size of the French army (variable 1) with its color indicating direction of march: light going in, dark going out (variable 2). There are some numbers along its

length to add detailed quantitative information to the general picture (starting at 422,000 and ending at 1,000). The location of the army is shown on the figurative map (variables 3 and 4), and the temperature and date for the return trip is shown on the bottom (variables 5 and 6). The picture is clear, and after a short while of studying it one can almost see the bedraggled remnants of the French army trudging through the horror of the Russian winter, leaving behind their fallen comrades on the frozen countryside. It has been said (Tufte, 1982) that when this graph was first shown publicly it brought tears to the eyes of the Frenchmen who saw it. A touching tribute to an evocative display of statistical data.

Insert Figure 36 about here

Summing Up

Although the tone of this presentation tended to be light, and pointed in the wrong direction, the aim is serious. There are many paths that one can follow which will deteriorate the quality of our data displays; the 12 'rules' that we described were only the beginning. Nevertheless, they point clearly toward an outlook that provides many hints for good display. The measures of display described interlock. The data density cannot be high if the graph is cluttered with chartjunk; the data-ink ratio grows with the amount of data displayed; the lie factor manifests itself most frequently when additional dimensions or worthless metaphors are included. Thus, the rules for good display are quite simple. Examine the data carefully enough to know what they have to say, and then let them say it with a minimum of adornment. Do this while following reasonable regularity practices in the depiction of scale and label clearly and fully. Last, and perhaps most important, spend some

time looking at the work of the masters of the craft. An hour spent with Playfair or Minard will not only profit your graphical expertise but also be enjoyable. Tukey's EDA contains 236 graphs and little chartjunk. The work of Francis Walker is clear and concise, and it is truly a mystery why current Census documents do not utilize more the schema developed a century and more ago.

TABLE 1

Data Density in Twenty Randomly Selected Statistical Graphics for Some Recent Publications
(in number of numbers per square inch)

	<u>Median</u>	<u>Minimum</u>	<u>Maximum</u>
Nature	48.0	3.0	362.0
Journal of the Royal Statistical Society, B	27.3	4.1	114.5
Science	20.8	4.6	44.0
Wall Street Journal	19.2	3.1	154.4
Fortune	17.9	5.3	31.2
The Times (London)	17.8	1.9	122.3
JASA	16.9	4.0	166.5
Le Monde	8.1	0.5	16.6
Psychological Bulletin	7.5	0.8	73.7
Journal of the American Medical Association	7.4	0.9	38.7
The New York Times	7.4	1.0	13.2
Newsweek	5.8	0.5	13.0
Scientific American	5.1	0.5	69.4
Statistical Abstract of the United States, 1979	4.6	2.0	23.0
Social Indicators III	0.6	0.2	1.8
Pravda	0.2	0.1	1.0

Source: Tufte, 1982

TABLE 2

Optimal Selection from a Finite Sequence with Sampling Cost

$g(X_{s+r-1}) = bR(X_{s+r-1}) + a, \text{ if } S = s$ $0, \text{ otherwise}$						
b/c = 10.0			100.0		1000.0	
N	r*	(G _N (r*) - a)/c	r*	(G _N (r*) - a)/c	r*	(G _N (r*) - a)/c
3	2	.20000	2	2.22500	2	22.47499
4	2	.26333	2	2.88833	2	29.13832
5	2	.32333	3	3.54167	3	35.79166
6	3	.38267	3	4.23767	3	42.78764
7	3	.44600	3	4.90100	3	49.45097
8	3	.50743	4	5.57650	4	56.33005
9	3	.56743	4	6.26025	4	63.20129
10	4	.62948	4	6.92358	4	69.86462

Source: Dhariyal & Dudewicz, 1981

TABLE 3

Optimal Selection from a Finite Sequence with Sampling Cost (revised)

$g(X_{s+r-1}) = bR(X_{s+r-1}) + a, \text{ if } S = s$ $0, \text{ otherwise}$							
N	b/c=10.0		b/c=100		b/c=1000		
	r*	G	r*	G	r*	G	
3	2	.2	2	2.2	2	22	
4	2	.3	2	2.9	2	29	
5	2	.3	3	3.5	3	36	
6	3	.4	3	4.2	3	43	
7	3	.4	3	4.9	3	49	
8	3	.5	4	5.6	4	56	
9	3	.6	4	6.3	4	63	
10	4	.6	4	6.9	4	70	

Figure Captions

Figure

- 1 An example of a low density graph (from SI3 [ddi=.3])
- 2 A low density graph (from Friedman and Rafsky, 1981 [adi=.5])
- 3 A high density graph (from The New York Times, 1981 [ddi=65])
- 4 A low density graph (from The Washington Post, 1978) with chart-junk to fill in the space (ddi=.2)
- 5 Hiding the data in the grid (from Kelly, Ayres and Bowen, 1967)
- 6 A contentless graph (from Crotty, 1970 [ddi=0])
- 7 A redone example of the data from Figure 5 (from Tufte, 1970)
- 8 Hiding the data in the scale (from SI3)
- 9 Expanding the scale and showing the data in Figure 8
- 10 The White House showing the influence of the Pravda School of Ordinal Graphics (from The New York Times, 1981)
- 11 Ignoring the visual metaphor (from The New York Times, 1978)
- 12 Reversing the metaphor in mid-graph while changing scales on both axes (from The New York Times, June 14, 1981)
- 13 Figure 12 redone with consistent scale and visual metaphor
- 14 A plot on the same topic done well two centuries earlier (from Play-fair, 1786)
- 15 An example of how to goose up the effect by squaring the eyeball (from The Washington Post, 1978)
- 16 The data in Figure 15 as an unadorned line chart (from Wainer, 1980)
- 17 Cubing the visual effect and choosing origin to yield a near record lie factor of over 131,000% (from The Washington Post)

Figure

- 18 Data from Figure 17 redone without tricks (from Wainer, 1980)
- 19 Hiding the effect by the careful choice of scale and origin
(from The Washington Post)
- 20 Regraph of data from Figure 19 with expanded scale, different starting point, and previous year's average added for context (from Wainer, 1980)
- 21 Changing scale in mid-axis to make large differences small (from the New York Post, 1981)
- 22 Changing scale in mid-axis to make exponential growth linear (from The Washington Post)
- 23 Data from Figure 22 redone with linear scale (from Wainer, 1980)
- 24 Emphasizing the trivial: hiding the main effect of sex differences in income through the vertical placement of plots (from SI3)
- 25 Figure 24 redone with two plots horizontally opposed showing sex differences more clearly
- 26 Figure 24 redone with the large main effects emphasized and the small one (time trends) suppressed
- 27 Jiggling the baseline makes comparisons more difficult (from the Handbook of Agricultural Charts)
- 28 An alternative version of Figure 27 with a straight line used as the basis of comparison.
- 29 Austria First! Obscuring the data structure by alphabetizing the plot (from SI3)
- 30 Ordering and spacing the data from Figure 29 as a stem-and-leaf diagram provides insights invisible previously

Figure

- 31 Ordering the bar chart by the data tells the tale a bit more clearly
 (from SI3)
- 32 Mixing a changed metaphor with a tiny label reverses the meaning of
 the data (from The New York Times, 1978)
- 33 Figure 32 redrawn with 1978 data placed on a comparable basis (from
 Wainer, 1980)
- 34 An extra dimension confuses even the grapher (from The Washington
 Post, 1979)
- 35 Data from Figure 34 redrawn simply (from Wainer, 1980)
- 36 Minard's (1824) graph of the French Army's ill-fated foray into Russia.
 A candidate for the title of "World's Champion Graph."

References

- Bertin, J. Semiologie Graphique. The Hague Mouton-Gautier (2nd ed.), 1973.
- Bertin, J. La Graphique Et Le Traitement Graphique De L'Information. France: Flammarion, 1977.
- Bertin, J. Graphics and the Graphical Analysis of Data. Translation, W. Berg, Technical Editor, H. Wainer. Berlin: DeGruyter, 1981.
- Cox, D. R. Some remarks on the role in statistics of graphical methods. Applied Statistics, 1978, 27, 4-9.
- Crotty, W. J. (Ed.). Public Opinion and Politics: A Reader. New York: Holt, Rinehart and Winston, 1970, 364.
- Dhariyal, I. D., & Dudewicz, E. J. Optimal Selection from a Finite Sequence with Sampling Cost. Journal of the American Statistical Association, 1981, 76:376, pp. 952-959.
- Ehrenberg, A. S. C. Rudiments of numeracy. Journal of the Royal Statistical Society, A, 1977, 140:277-297.
- Friedman, J. H., & Rafsky, L. C. Graphics for the Multivariate Two-Sample Problem. Journal of the American Statistical Association, 1981, 76: 374, 277-287.
- Joint Committee on Standards for Graphic Presentation, Preliminary Report, Journal of the American Statistical Association, 1915, 14, 790-797.
- Kelley, S., Ayres, R. E., & Bowen, W. G. Registration and Voting: Putting First Things First. American Political Science Review, 1967, 61:371.
- Mayr, G. von. Gutachten uber die Anwendung der graphischen und geographischen Method in der Statistik, Munich, 1874.
- Minard, C. J. Tableaus Graphiques et Cartes Figuratives de M. Minard, 1845-1869.

- Playfair, W. The Commercial and Political Atlas, London, 1786.
- Schmid, C. F. Handbook of Graphic Presentation. New York: Ronald, 1954.
- Schmid, C. F., & Schmid, S. E. Handbook of Graphic Presentation (2nd ed.). New York: Wiley, 1979.
- Tufte, E. R. (Ed.). The Quantitative Analysis of Social Problems. Reading, Massachusetts: Addison-Wesley, 1970.
- Tufte, E. R. Improving data display. Department of Statistics, University of Chicago, 1977.
- Tufte, E. R. The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press, 1982.
- Tukey, J. W. Exploratory Data Analysis. Reading, Massachusetts: Addison-Wesley, 1977.
- Wainer, H. Making newspaper graphs fit to print. In Bouma, Kolers & Wrolsted (Eds.), Processing of Visible Language, (Vol. 2). New York: Plenum, 1980.
- Wainer, H. How are we doing? A review of Social Indicators III. Journal of the American Statistical Association, in press.
- Wainer, H., & Thissen, D. Graphical data analysis. Annual Review of Psychology, 1981, 32:191-241.
- Walker, F. A. Statistical Atlas of the United States Based on the Results of the Ninth Census. Bureau of the Census: Washington, D.C., 1894.

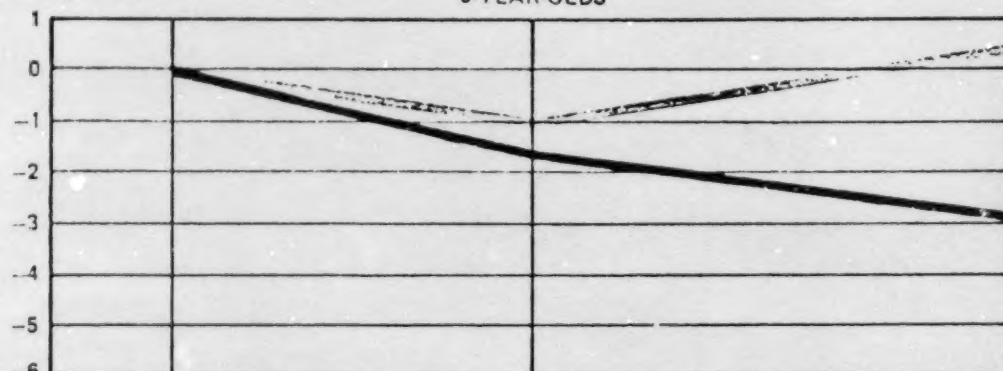
Figure 1

Chart 6/18. **Change in Science Achievement of 9-, 13-, and 17-Year-Olds, by Type of Exercise: 1969-1977**

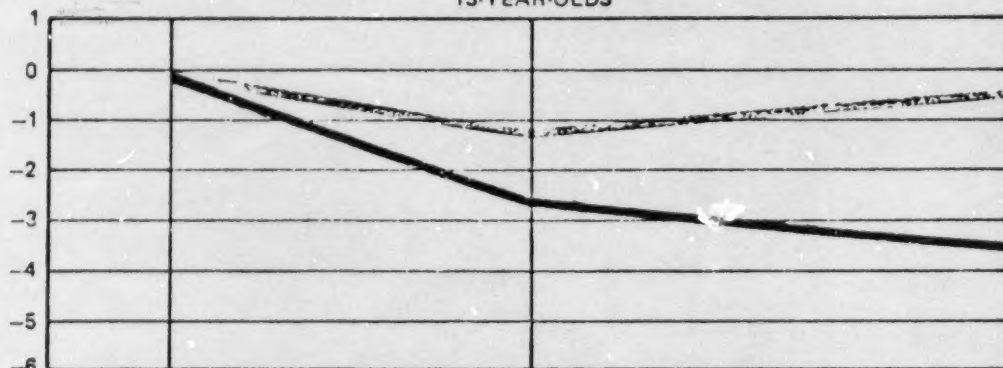
Biological science
Physical science

Change in percent correct

9-YEAR-OLDS



13-YEAR-OLDS



17-YEAR-OLDS

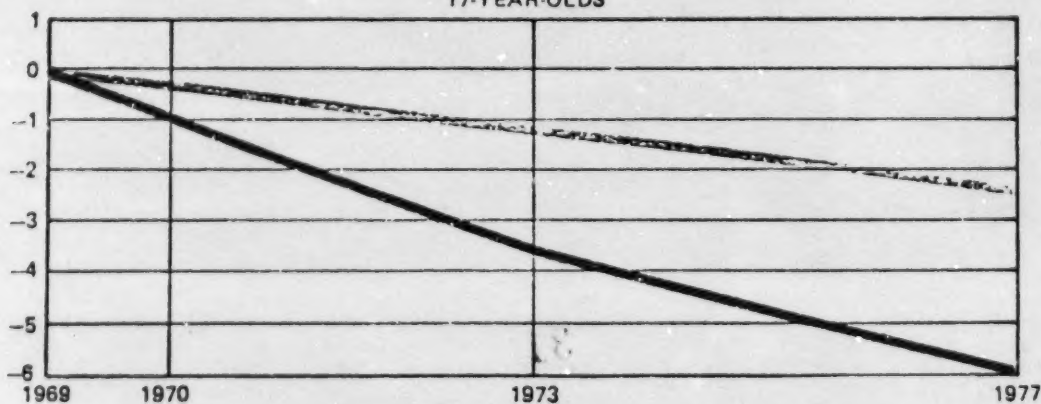
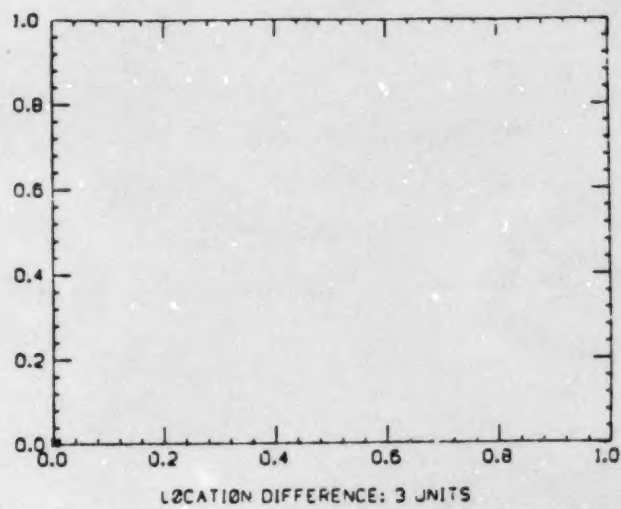


Figure 2



NEW YORK CITY'S WEATHER FOR 1981

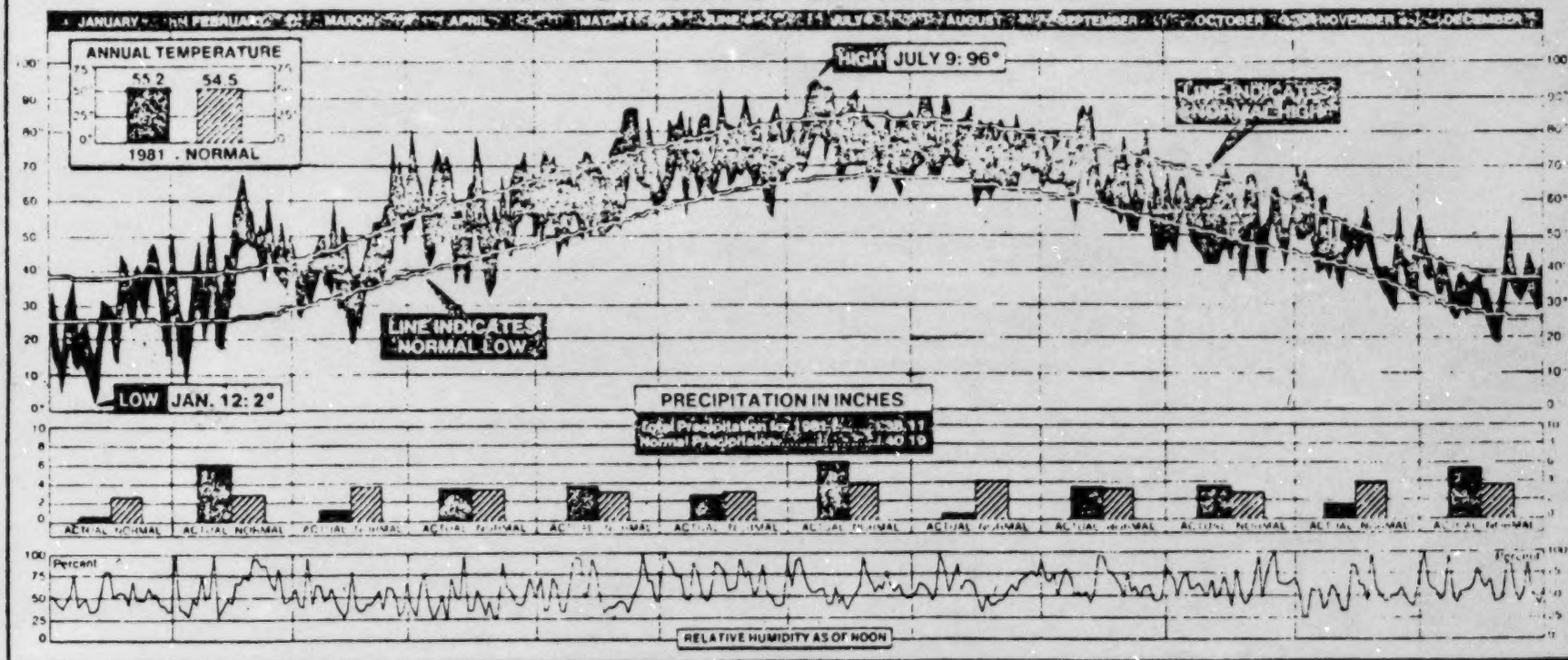


Figure 4

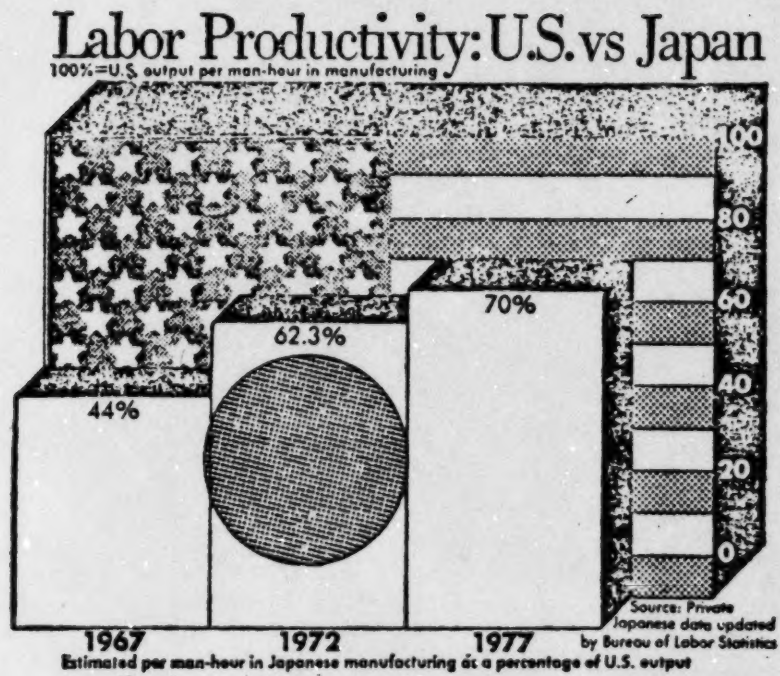
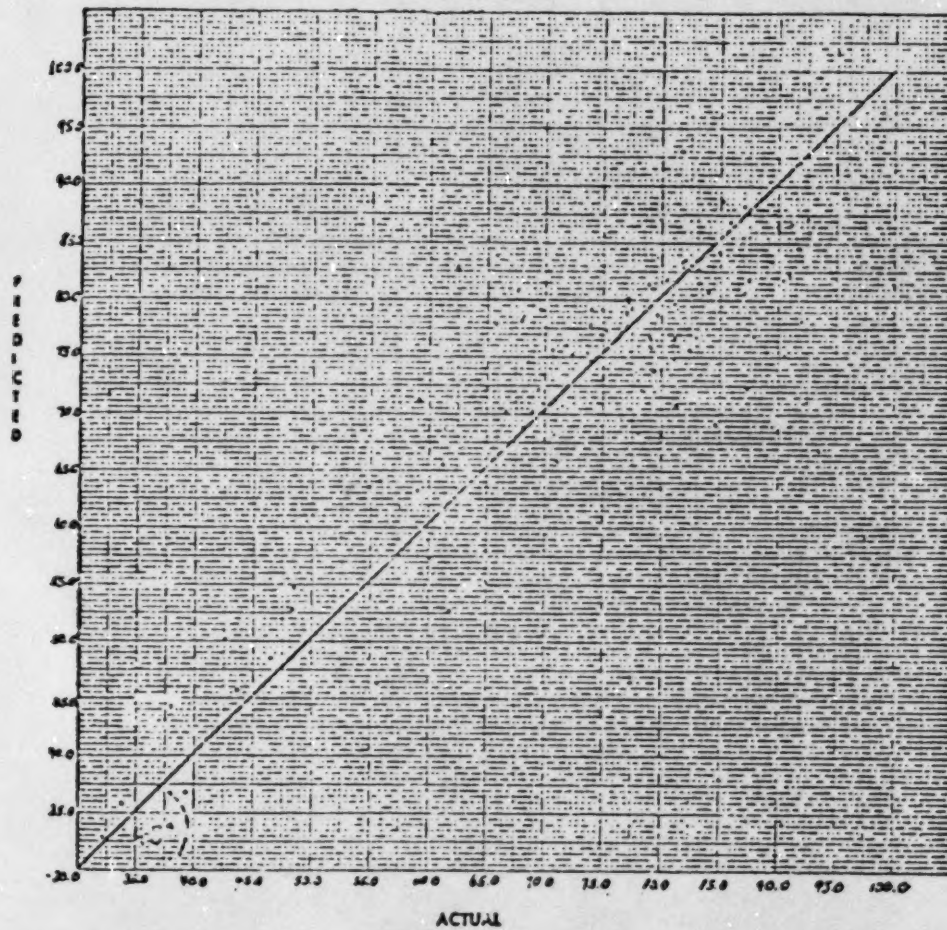


Figure 5

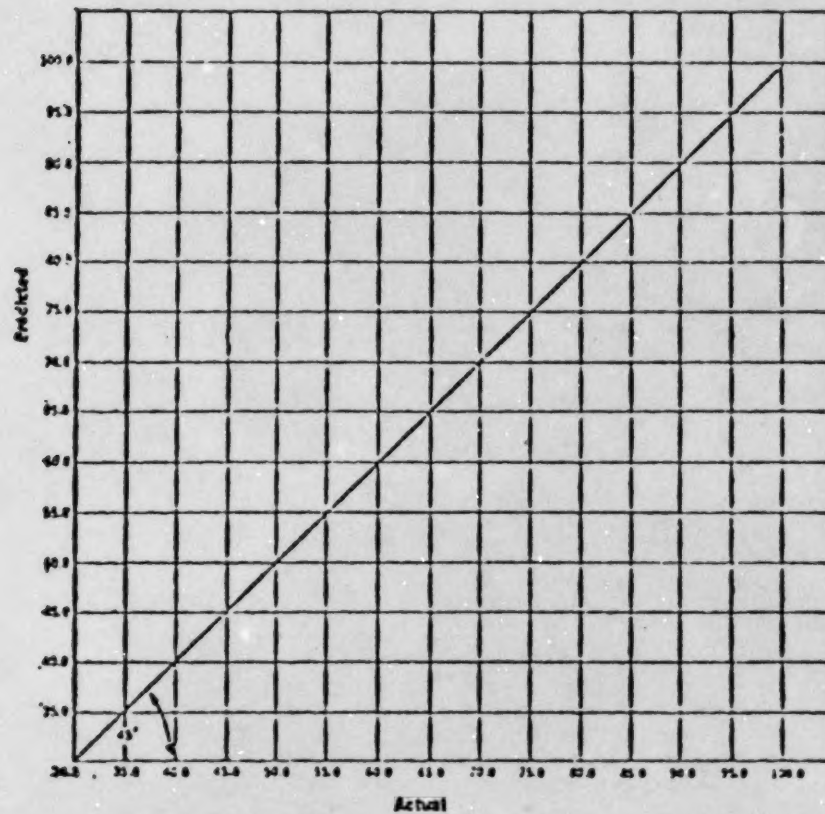
FIG. 1. Relationship of Actual Rates of Registration to Predicted Rates
(104 cities 1960).



Source: Kelley, S., Ayres, R. E., and Bowen, W. G. (1967) Registration and Voting: Putting First Things First. American Political Science Review, 61, p. 371.

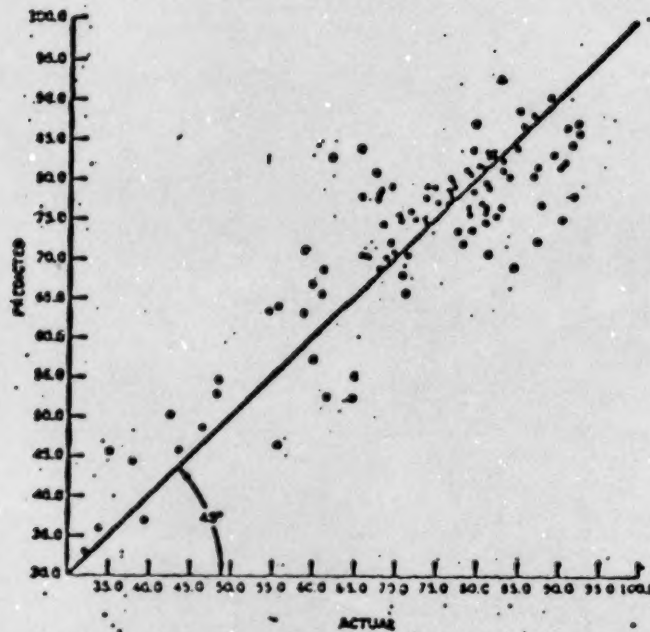
Figure 6

Figure 19.1 Relationship of Actual Rates of Registration to Predicted Rates
(104 cities, 1960)



Source: Crotty, W. J. (ed.) Public Opinion and Politics: A Reader
New York: Holt, Rinehart and Winston, 1970. P. 364.

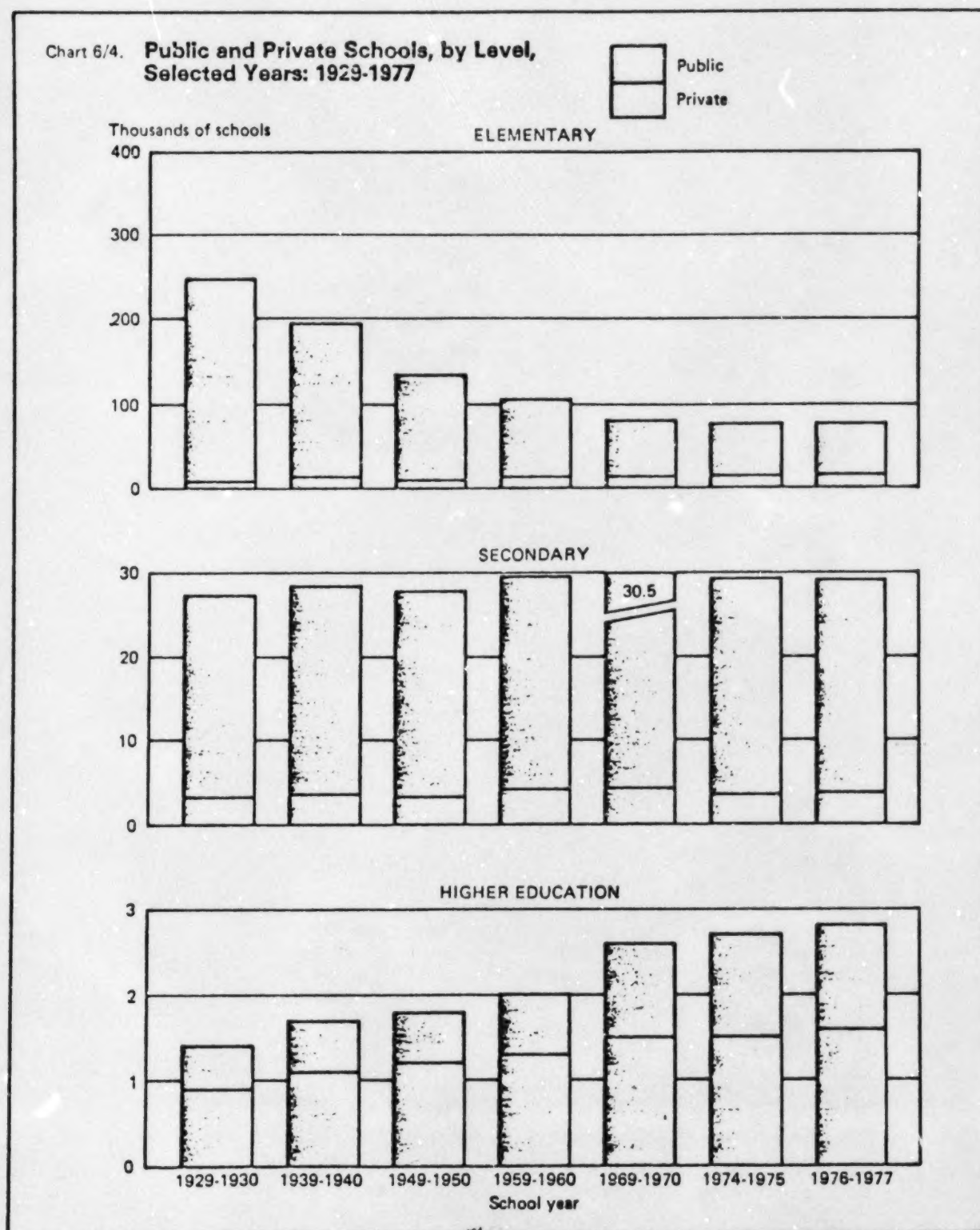
Figure 7



Relationship of Actual Rates of Registration to Predicted Rates (104 cities 1960).

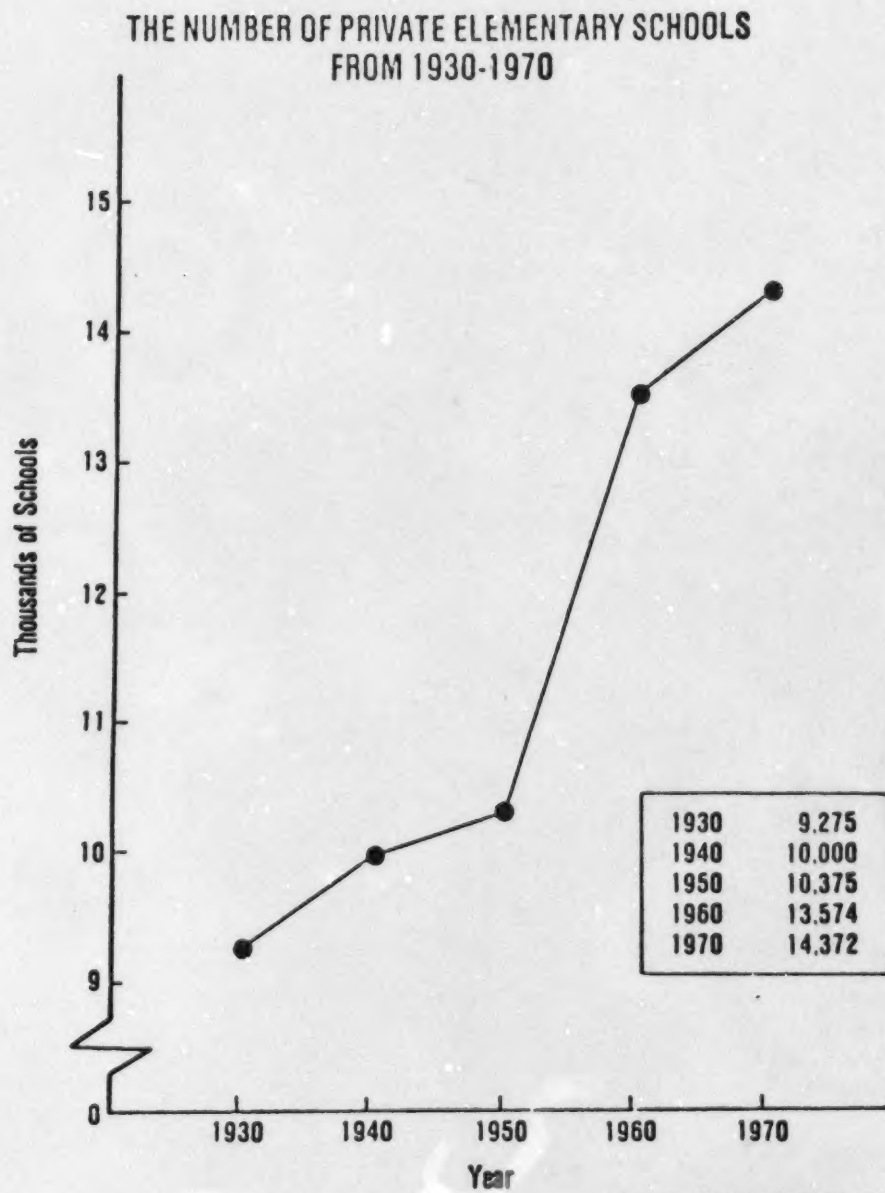
Source: Tufte, E. R. (ed.) The Quantitative Analysis of Social Problems. Reading, Massachusetts: Addison-Wesley, 1970.
P. 267.

Figure 8



Source: See table 6/4.

Figure 9



Source: Social Indicators III,
Table 8/4

OTHER BUSINESS

That Numberless Presidential Chart

"The lines on these charts say a lot about who's really fighting for whom."

— Ronald Reagan

Maybe, but with charts, reality tends to be in the eye of the artist.

The one President Reagan was showing off with such pride Monday night on national television showed two widely diverging lines, the gap between them in shrieking red, that represented the Administration's estimate of how much more one would be paying in taxes under the House Democrats's bill than under the President's.

Missing: Any clue at all to content. There was no scale, in fact no numbers whatever on the chart, except for the years 1982 and 1986.

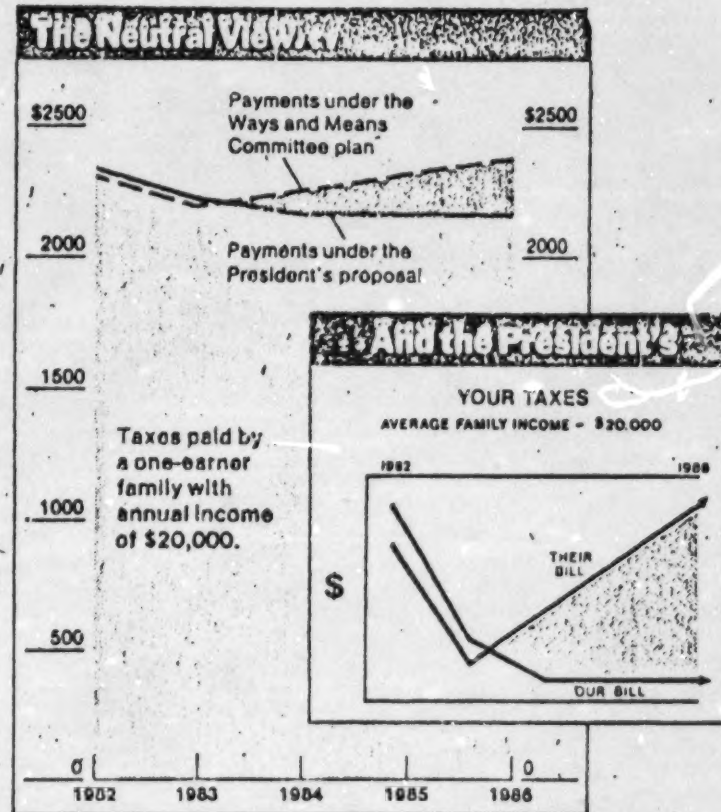
This, as any chartist will tell you, just won't do.

"It's unusual to have a graph without numbers," said David Garth, president of The Garth Groupings, political media consultants (and a well known Democrat), in a fine exercise in understatement. "Obviously he is using the graph because the numbers themselves are not that impressive."

"There are no laws in representational graphics," added Herbert Cytryn, creative director for the McGraw Hill Publishing Company, who went on to explain that "It is a very common practice not to start with a zero base if you want to magnify something out of proportion to your advantage. Technically it is not wrong. But it may be wrong morally."

"The average person," Mr. Cytryn concluded, "looks at it and is going to very impressed, but when they get the numbers..."

And so to the numbers. The President's chart, it turns out, shows that a



one-earner family with annual income of \$20,000 would, by 1986, be paying \$2,168 in taxes under his plan and \$2,385 under the House Ways and Means Committee's plan. (The first comprising a three-year, 25 percent tax cut and indexing payments to inflation in the third year, the second,

a two-year 15-percent tax cut and no indexing.)

Pointing out that big red gap, the President pounded in his point: the Democratic bill is better "if you only intend to live for two more years," he said. "This red space between the two lines is the tax money that will remain

in your pockets if our bill passes and it's the amount that will leave your pockets if their bill is passed."

"On the one hand you see a genuine and lasting commitment to the future of working Americans," the President stressed, without verbally presenting any numbers either. "On the other just another empty promise."

That's pretty strong stuff considering that the maximum difference one is talking about is \$217 in 1986.

The chartist's art, of course, does call for making the most dramatic statement consistent with the data. In this case, it was achieved by using the tightest possible scale, starting with an oddly unrounded \$2,150 and winding up at \$2,400. Using the purist's zero base (meaning starting at \$0 and going up to the first round number after the highest point in the chart) puts everything in rather a different perspective. That \$217 seems small beer in a total tax bill of \$2,385.

But what did happen to the numbers?

"The chart we sent over to the White House had all the numbers in it," said Marlin Fitzwater, a press officer at the Treasury Department.

"We tried it with numbers," David R. Gergen, senior White House spokesman, said "and found that it was very hard to read on television, so we took them off. We were trying to put a point across, not the absolute numbers."

"The President wanted to keep it as simple as possible," added Karna Small, deputy White House Press Secretary.

Finally, what does it all prove? "Reagan is a great communicator," said Gerald Rafshoon, former President Carter's media adviser.

Thomas L. Friedman

Figure 10

Figure 11

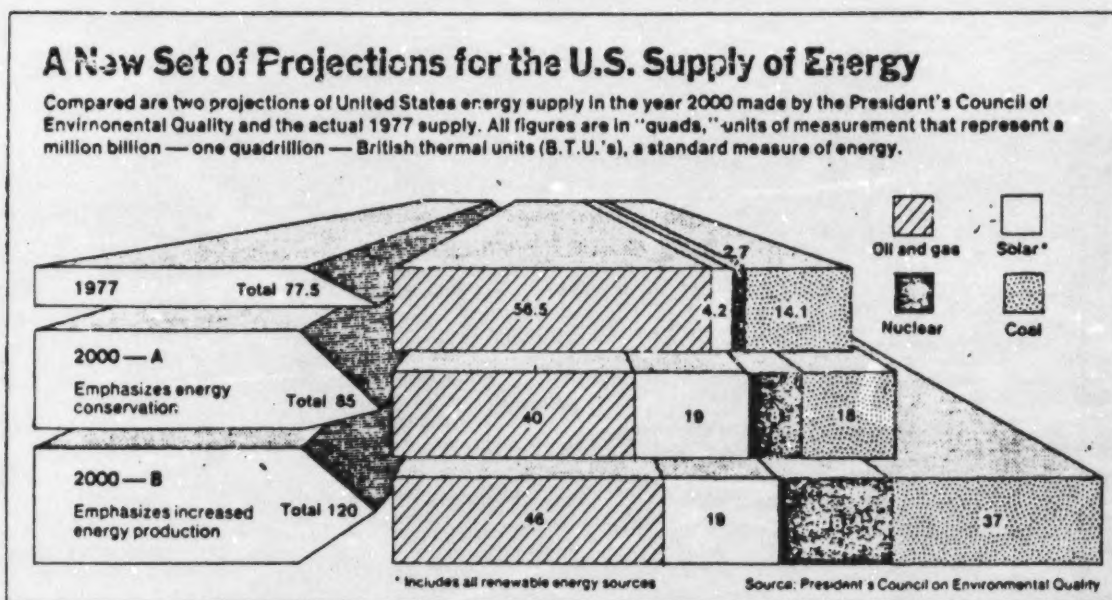


Figure 12

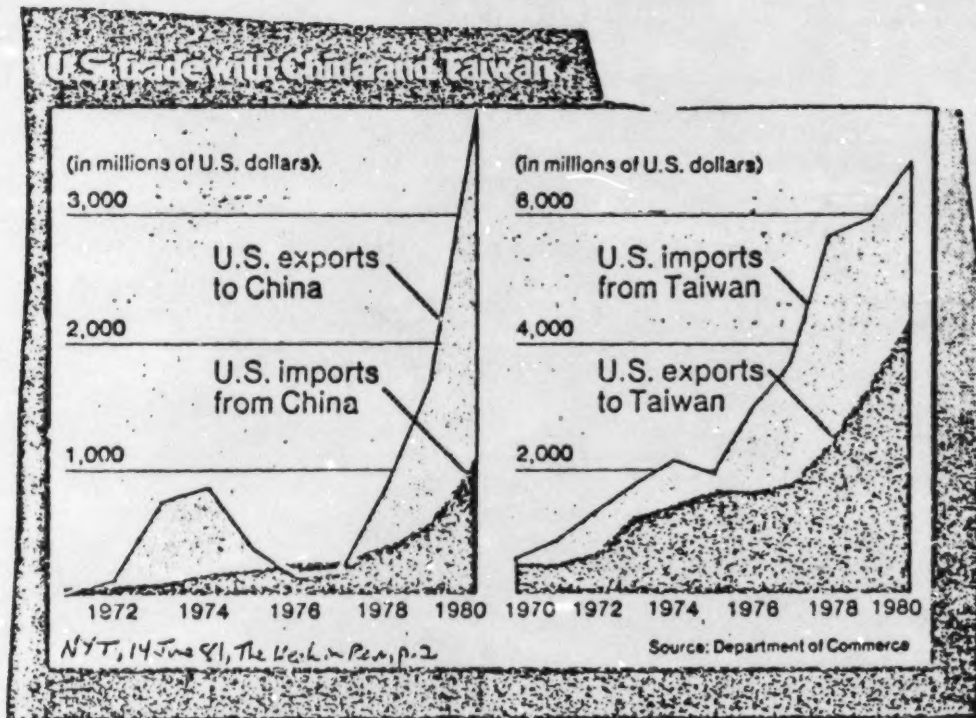
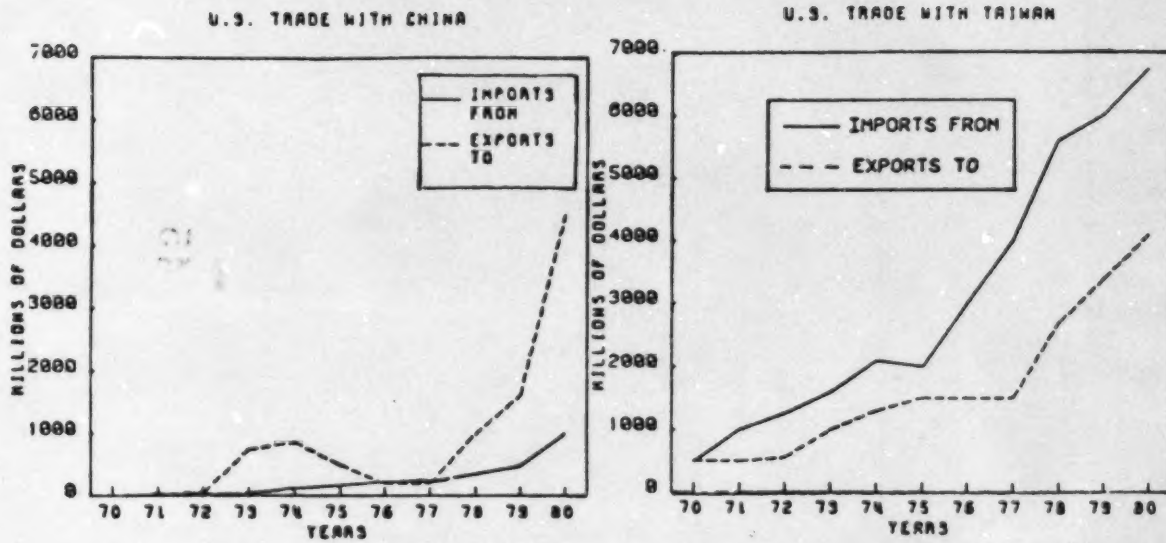


Figure 13



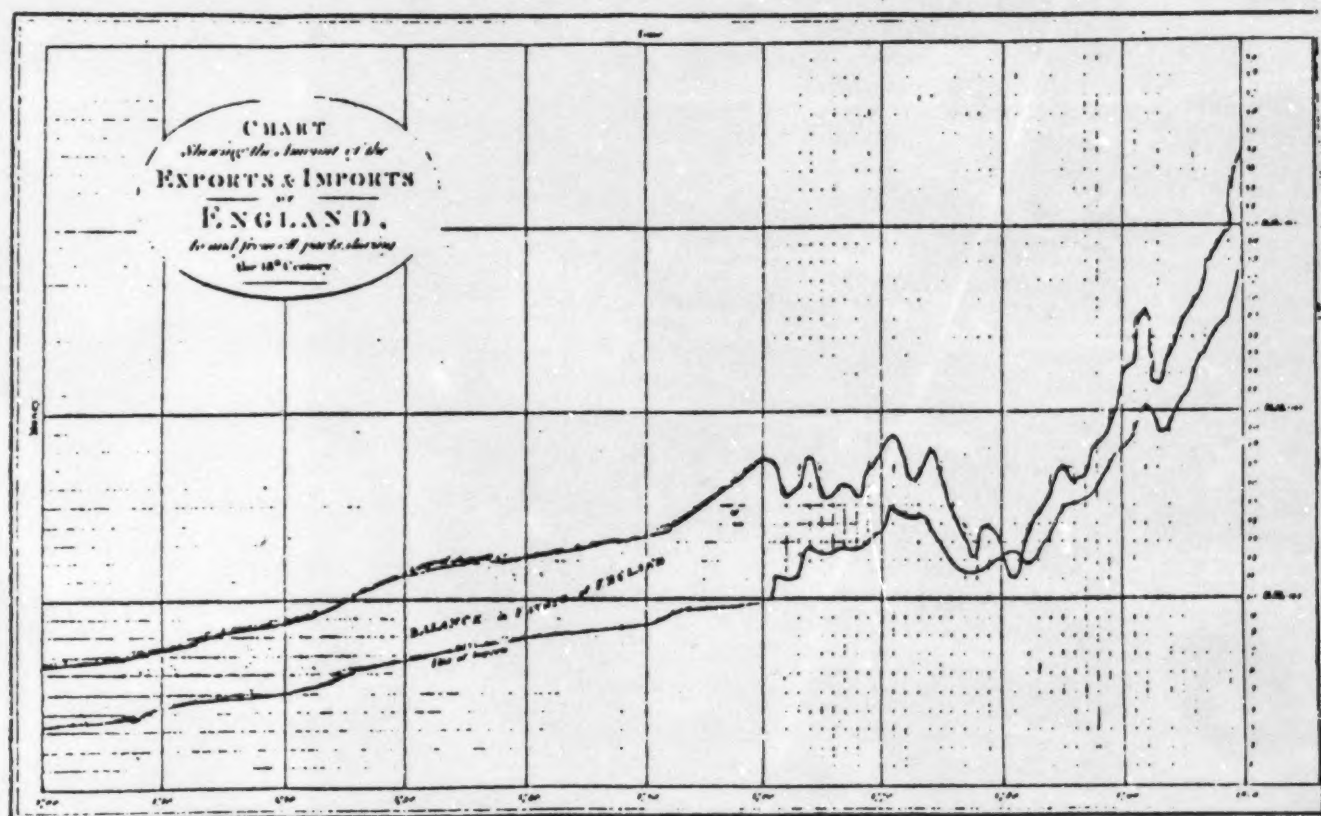


Figure 14

Figure 15



Figure 16

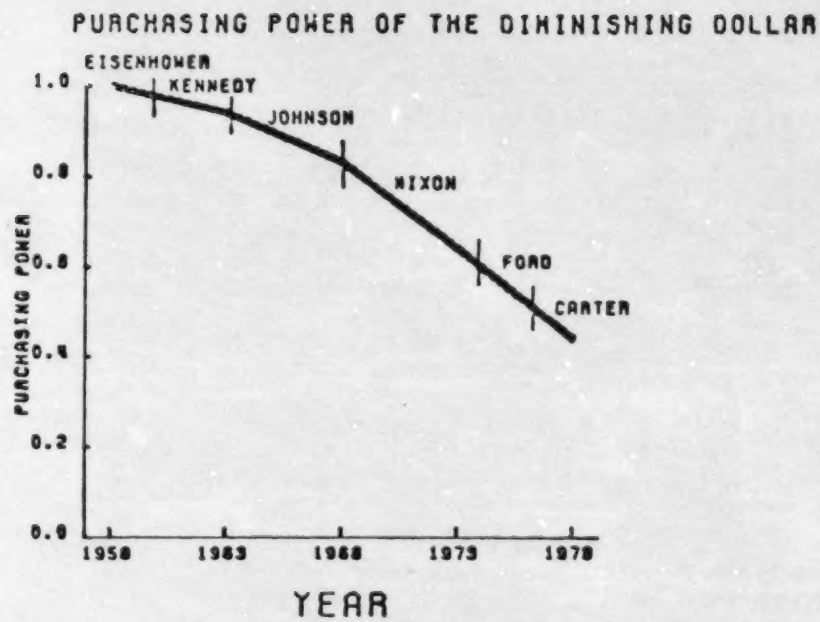
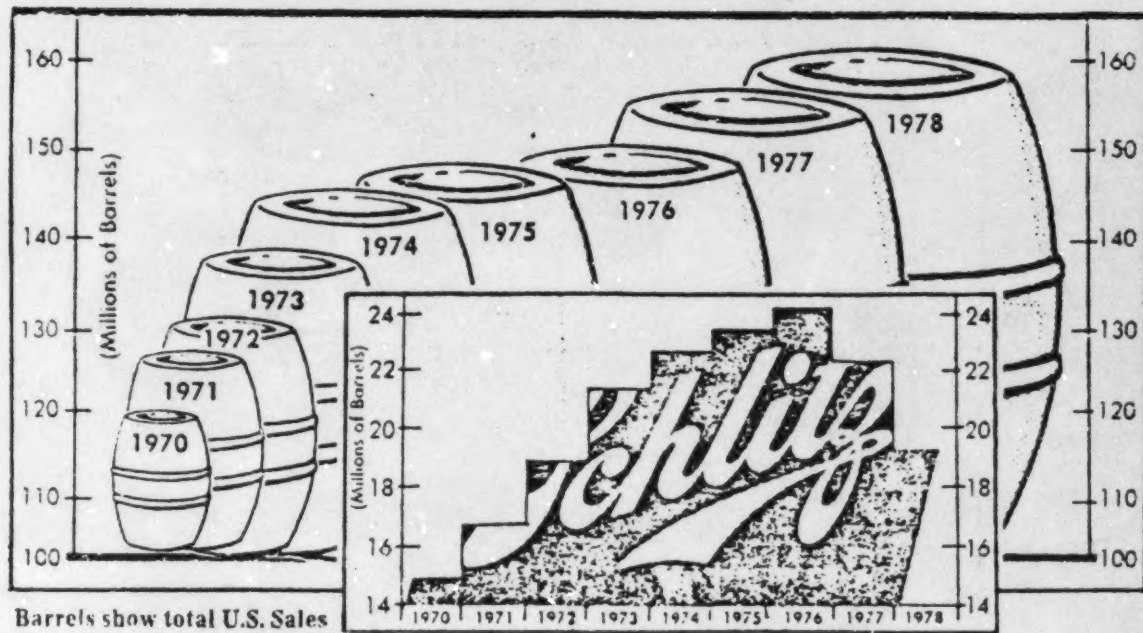


Figure 17

U.S. Beer Sales and Schlitz' Share



Barrels show total U.S. Sales

Figure 18

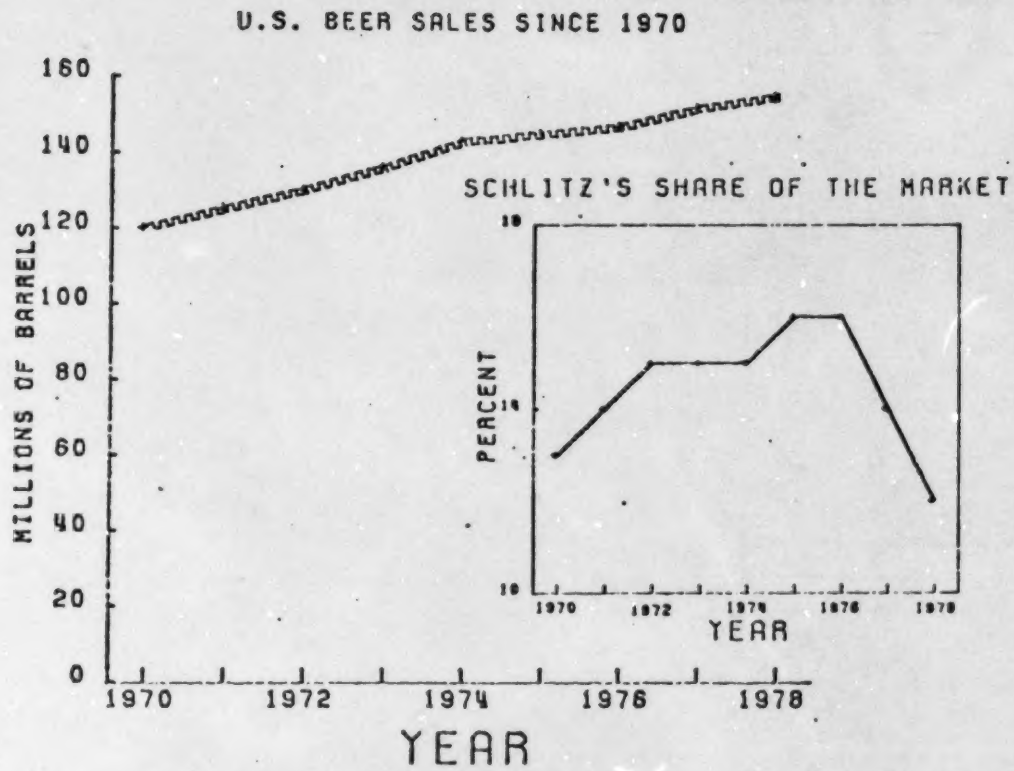


Figure 19

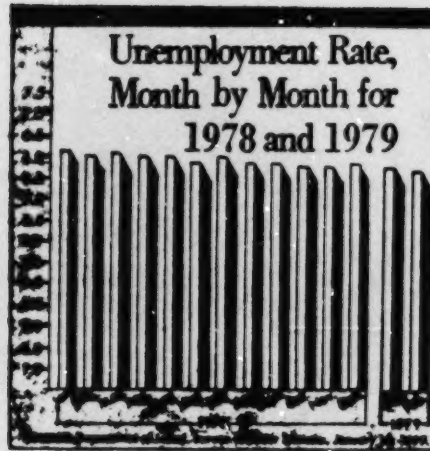
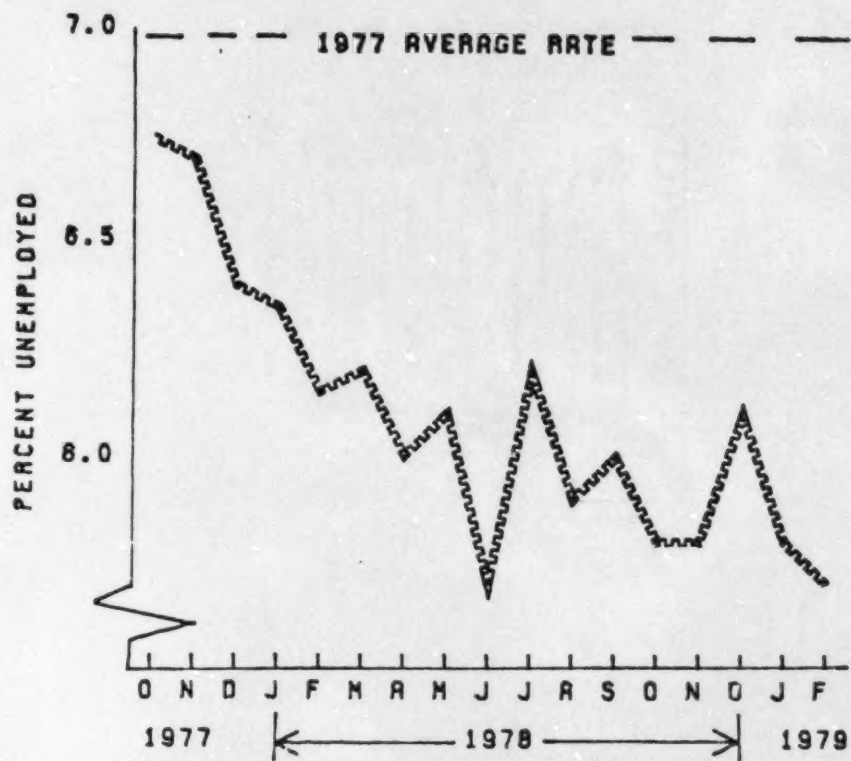


Figure 20



UP! UP! UP! UP! UP! UP! UP! UP! UP! UP! UP! UP! UP!

EVERY day more and more New Yorkers are turning to the NEW YORK POST. We know it, New Yorkers know it, and the latest official newspaper audit confirms it.

The Audit Bureau of Circulation reports that POST sales in the latest audit period, October 1980 to March 1981, rose sharply to 732,158 copies a day, 77,841 more than in the same period a year before, a very healthy 12 per cent increase.

In stark contrast, the sales of the Daily News tumbled by 61,000 to a new low of 1,491,556, and this despite the vastly expensive launch of its Tonight edition.

Indeed, since News executives still claim a sale of around 1,800,000 for their admittedly aging after

The soaraway Post — the daily paper New Yorkers trust

noon venture, the figures suggest that the morning sales of the News have dropped by a staggering 163,000 copies a day.

Not for nothing is the POST hailed as America's fastest growing newspaper. On the other hand, as the chart (based on official ABC figures) shows, the Daily News, with its steady and relentless drop

from over 2 million in 1974, has earned the title of America's fastest falling newspaper.

The POST also outperformed the Daily News on Saturdays. Sales of the Saturday POST increased by 59,000 copies, from 421,350 to 479,350, a solid 14 per cent.

The Sunday sales of the

News fell below the 2 million mark for the first time since World War II, from 2,262,601 to 1,995,702, an exceptionally severe drop of 266,899.

And to give you some idea why so many readers trust THE POST, read what our man Mike Marley predicted on the big fight seven hours before the bell rang: Doug 72

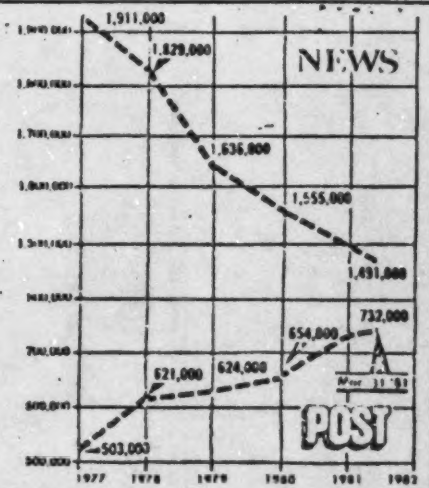


Figure 21

BEST COPY AVAILABLE

Figure 22

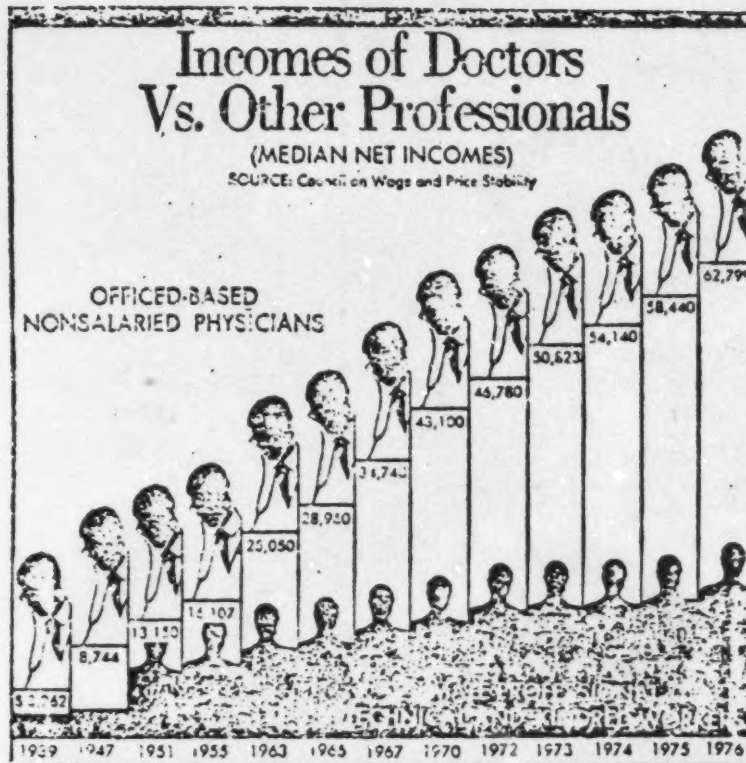


Figure 23

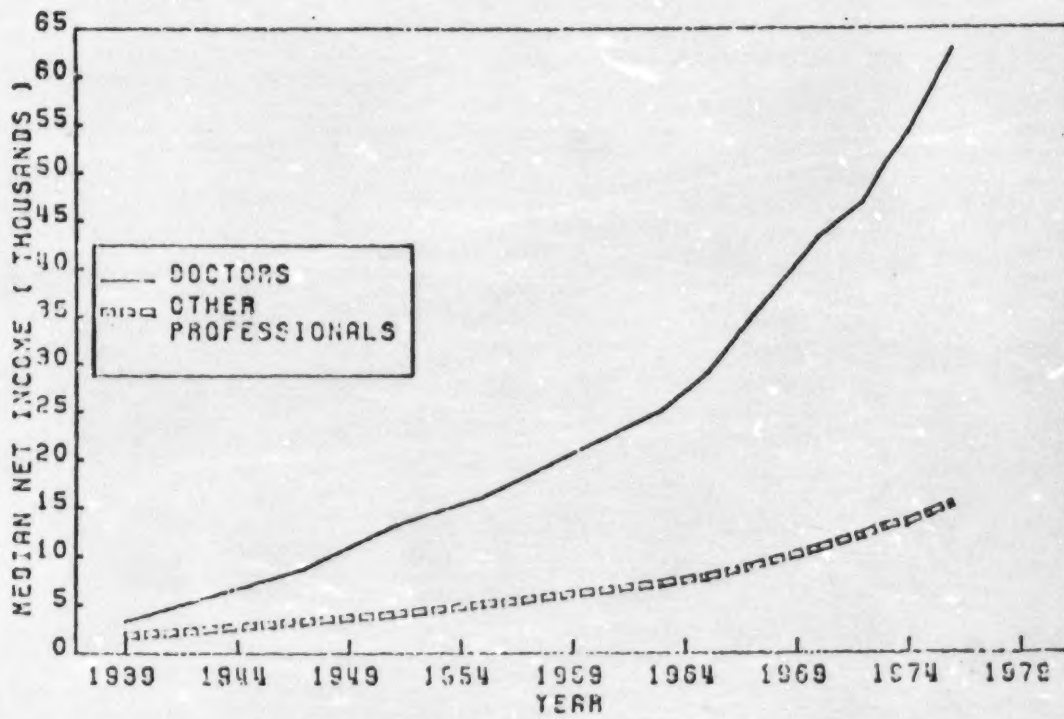
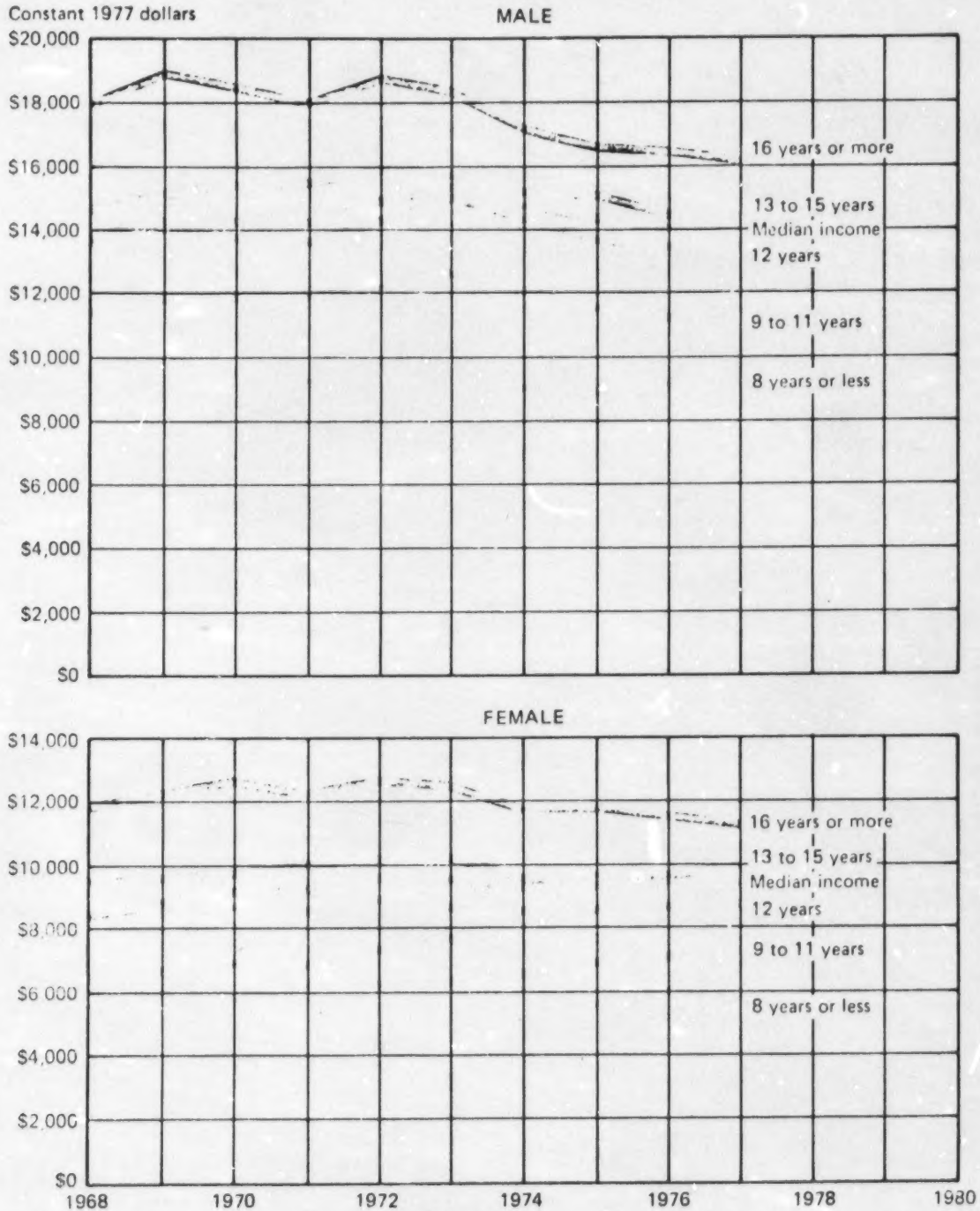


Figure 24

Chart 6/19. **Median Income of Year-Round, Full-Time Workers 25 to 34 Years Old, by Sex and Educational Attainment: 1968-1977**



**MEDIAN INCOME OF YEAR-ROUND FULL TIME WORKERS
25-34 YEARS OLD BY SEX AND EDUCATIONAL ATTAINMENT:
1968-1977**

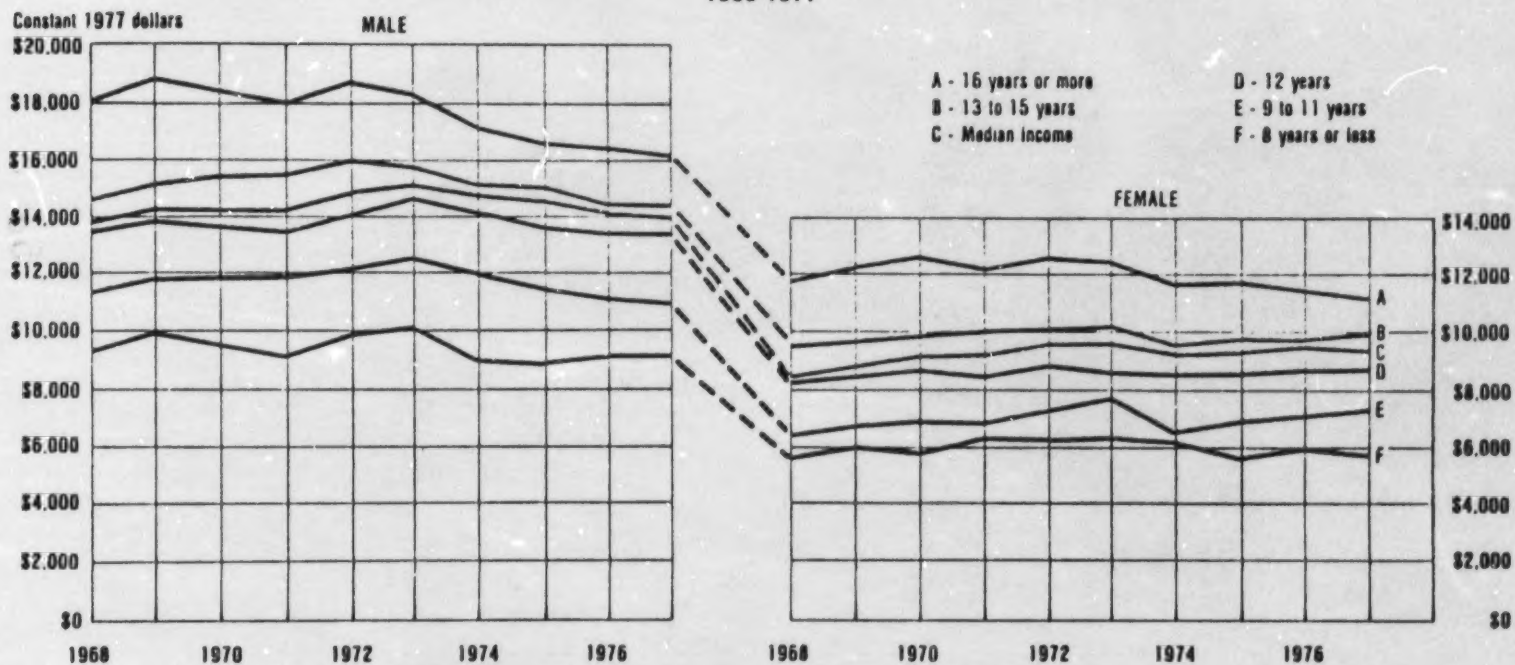
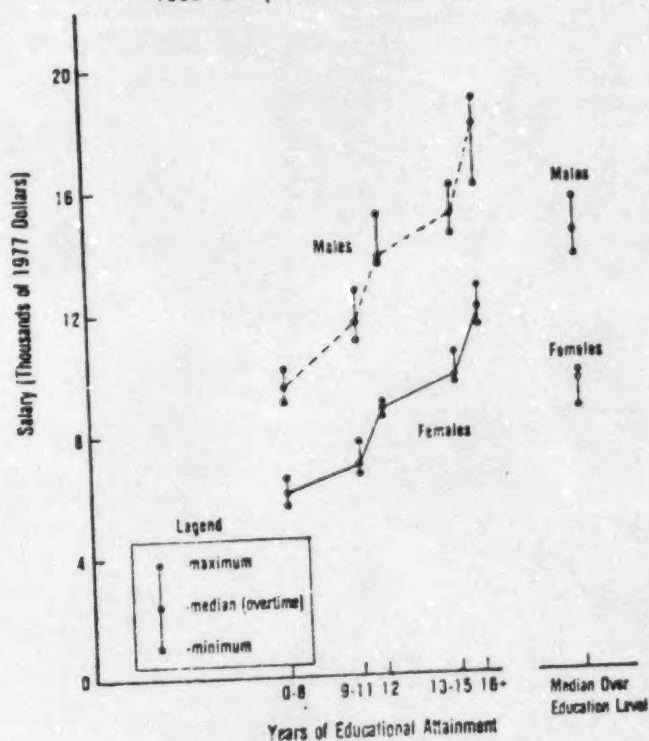


Figure 25

Social Indicator III
Table 6/19

Figure 26

MEDIAN INCOME OF YEAR-ROUND FULL TIME WORKERS
 25-34 YEARS OLD BY SEX AND EDUCATIONAL ATTAINMENT:
 1968-1977 (IN CONSTANT 1977 DOLLARS)



Source: Social Indicators III, 8/19

Figure 27

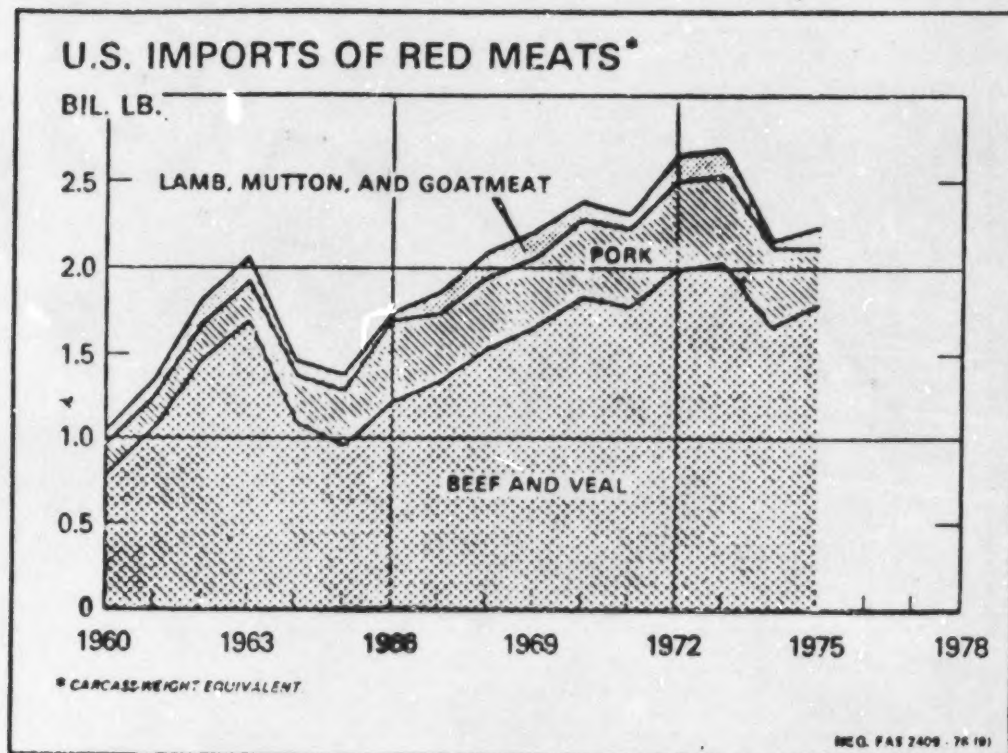


Figure 28

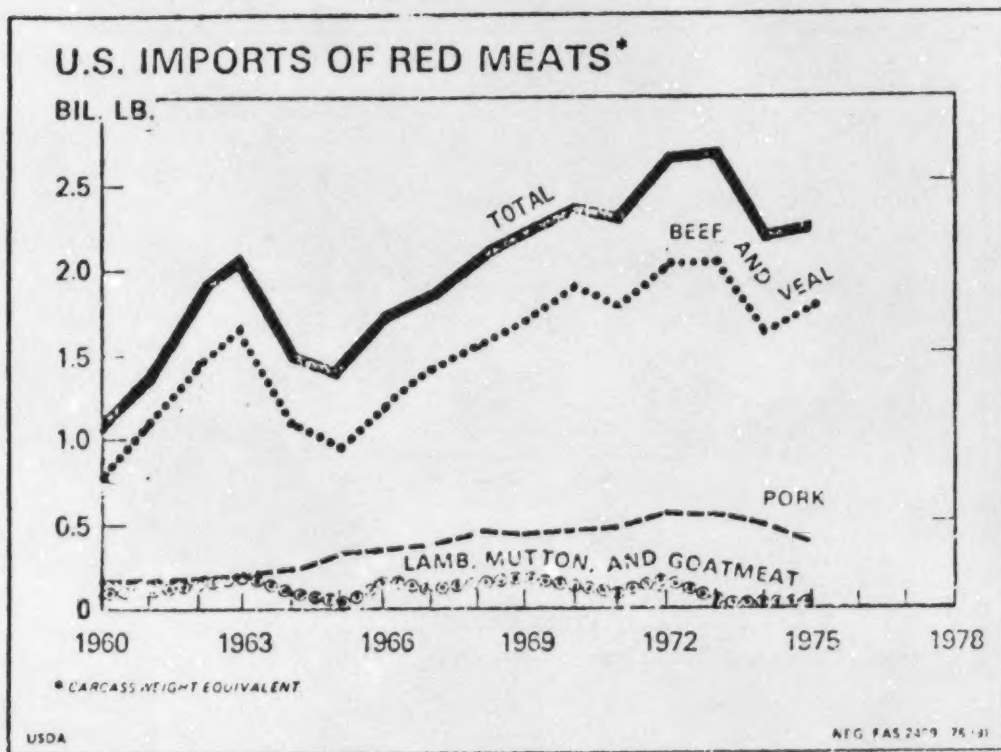


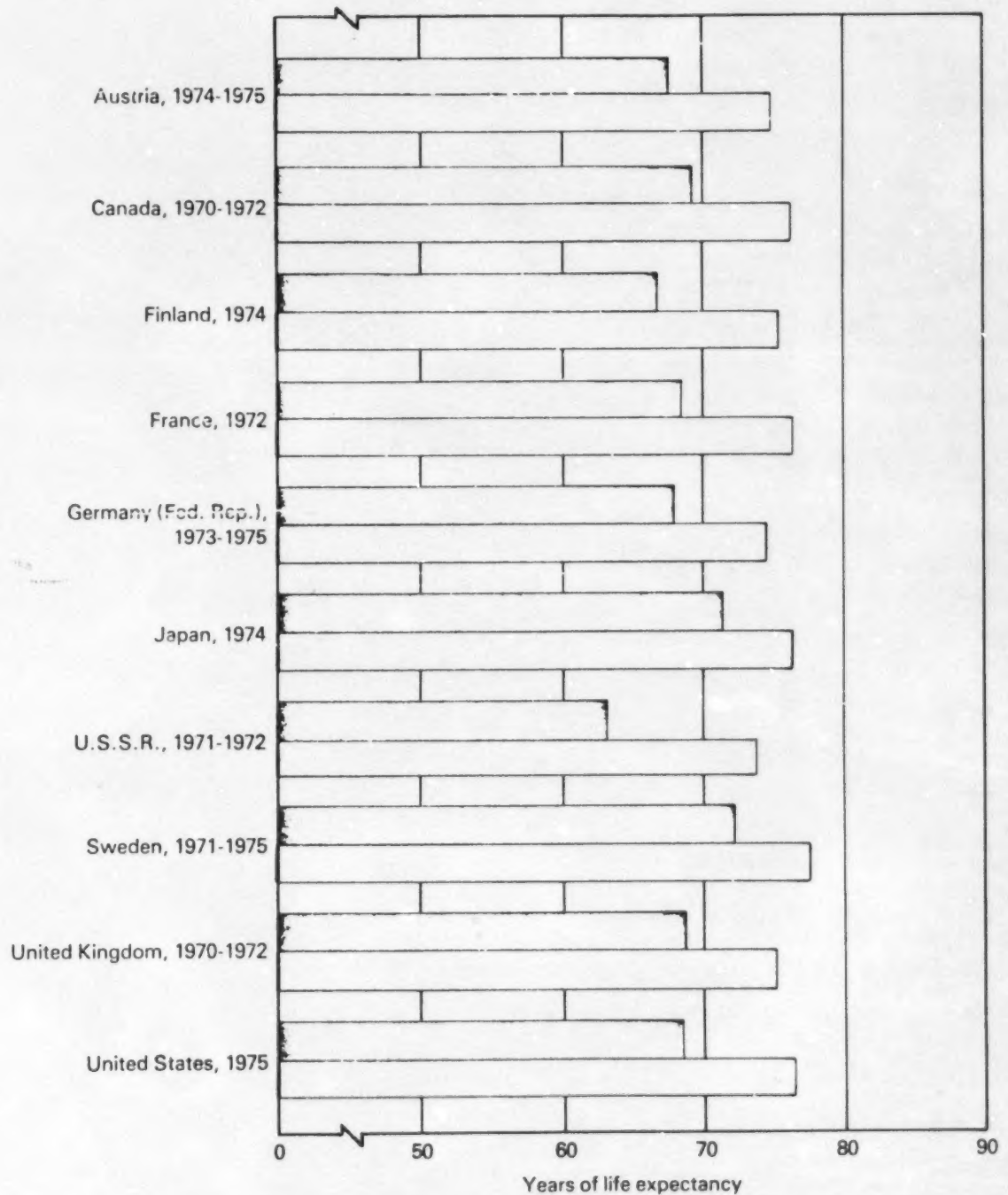
Figure 128

Source: Handbook of Agricultural Charts, U.S. Department of Agriculture, 1976, p. 93.
Chart Source: Original

Figure 29

Chart 2/29. **Life Expectancy at Birth, by Sex, Selected Countries, Most Recent Available Year: 1970-1975**

Male
Female



Source: See table 2/29.

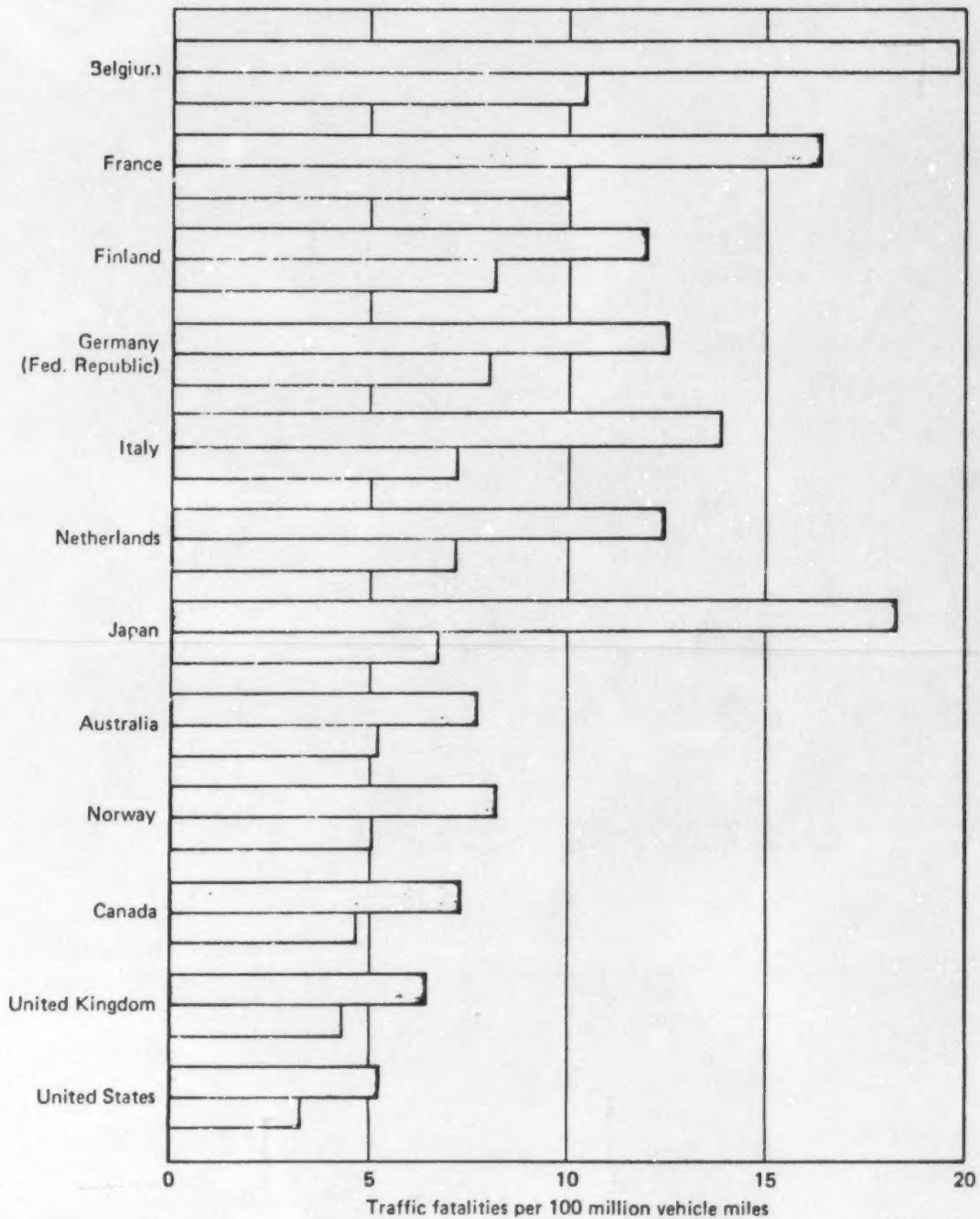
Figure 30

LIFE EXPECTANCY AT BIRTH, BY SEX,
MOST RECENT AVAILABLE YEAR

<u>WOMEN</u>	<u>YEARS</u>	<u>MEN</u>
SWEDEN	78	
	77	
FRANCE, US, JAPAN, CANADA	76	
FINLAND, AUSTRIA, UK	75	
USSR, GERMANY	74	
	73	
	72	SWEDEN
	71	JAPAN
	70	
	69	CANADA, UK, US, FRANCE
	68	GERMANY, AUSTRIA
	67	FINLAND
	66	
	65	
	64	
	63	USSR
	62	

SOURCE: SOCIAL INDICATORS III
TABLE 2/29

Chart 4/25. **Motor Vehicle Fatality Rates,
Selected Countries: 1969 and 1976**



Source: See table 4/25.

Figure 32



The New York Times / Aug. 2, 1978

Complex web of discount fares and airlines' telephone delays are raising travel agents' overhead, offsetting revenue gains from higher volume.

Figure 33

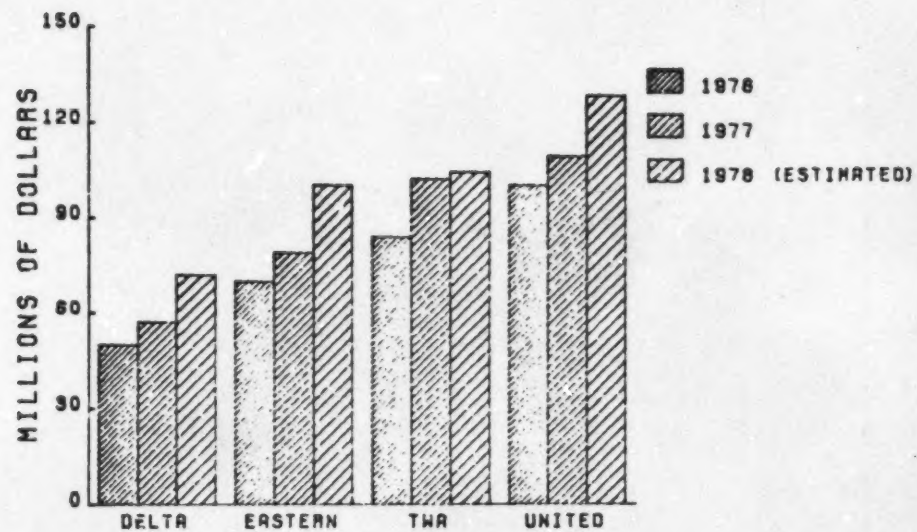


Figure 34

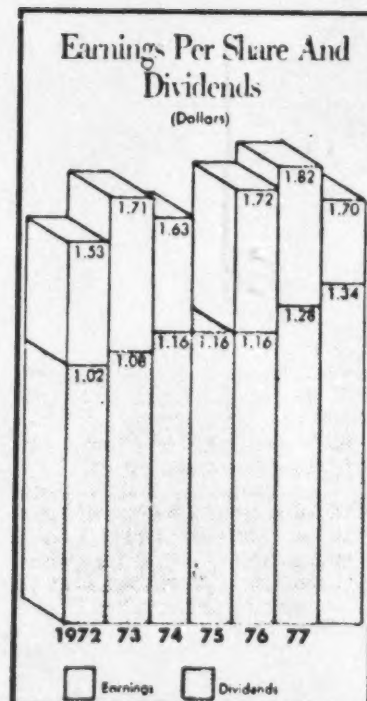
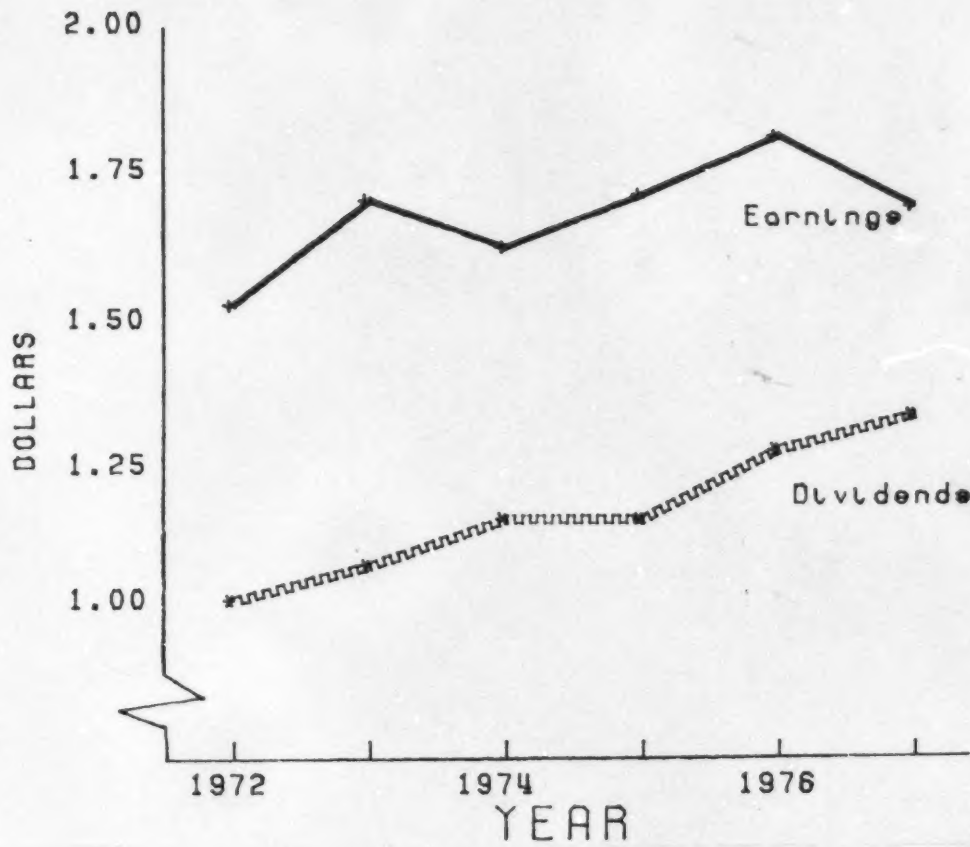


Figure 35



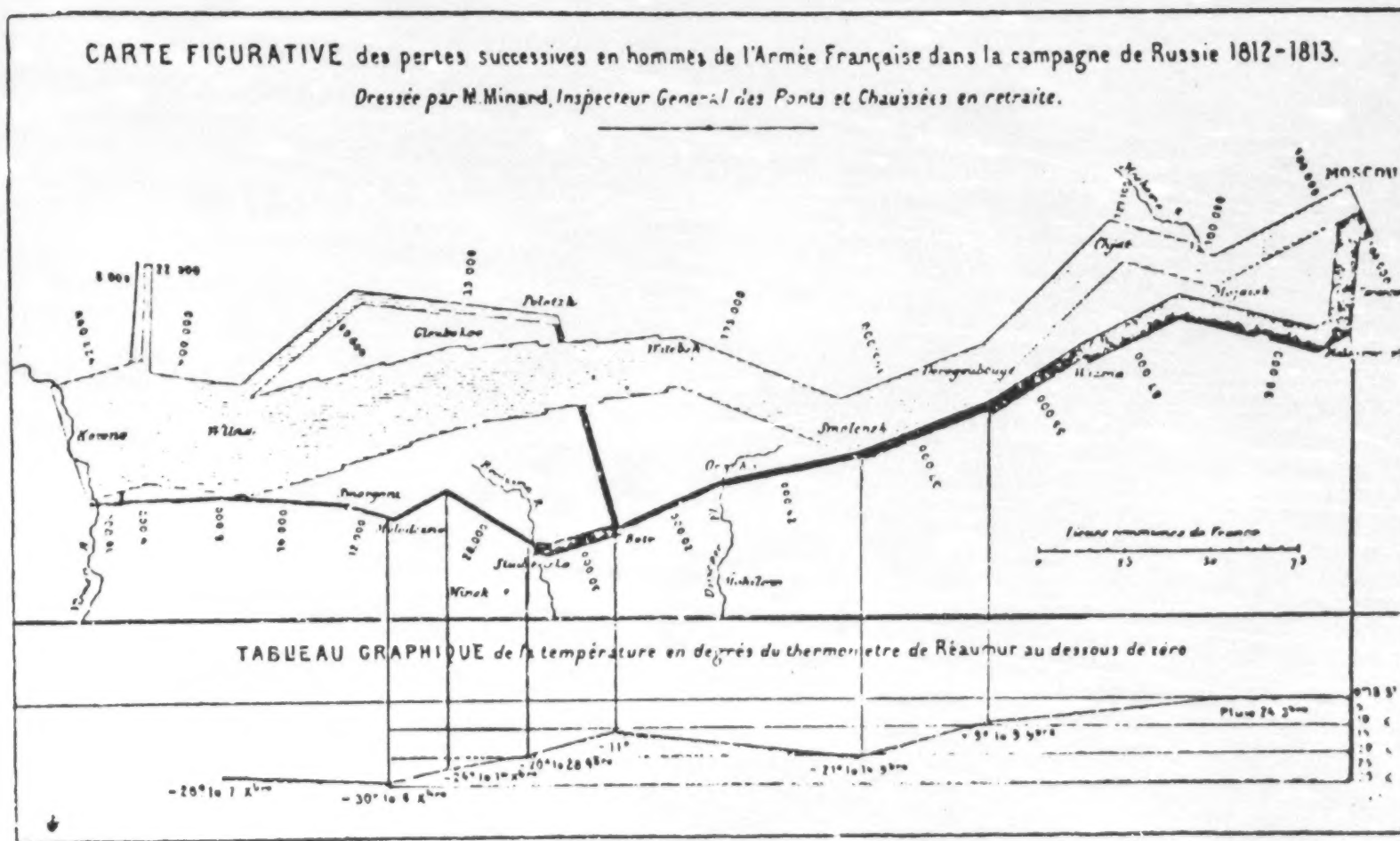


Figure 36

Russie boréale

X^{re} = Dniepr

Fig 37.

9^{bre} = Niémen

8^{bre} = Orjola

Charles Joseph Minard, *Des tableaux graphiques et des cartes figuratives* (Paris, 1861)

BEST COPY AVAILABLE